9 October 2006

Advanced information on Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel

Edmund Phelps’s Contributions to Macroeconomics
Edmund Phelps’s Contributions to Macroeconomics

The relationship between inflation and unemployment and the tradeoff between the welfare of current and future generations are key issues in macroeconomic research. They have a strong influence on the choice of macroeconomic policy. Both issues concern tradeoffs between different objectives. How should fiscal and monetary policy resolve the conflict between the goals of low inflation and low unemployment? How should society trade off consumption today against consumption in the future, i.e., how much should be saved in order to increase future consumption? Edmund Phelps has made major contributions to the analysis of both of these tradeoffs. In particular, he had the insight that also the balance between inflation and unemployment reflects a fundamentally intertemporal problem.

Phelps pointed out that current inflation depends not only on unemployment but also on inflation expectations. Such dependence is due to the fact that wages and prices are adjusted only infrequently. Consequently, when adjustments are made, they are based on inflation forecasts. Thus, the higher the anticipated rate of inflation, the higher the unemployment required to reach a certain actual inflation rate. Phelps formulated the so-called expectations-augmented Phillips curve. The intertemporal perspective implies that current inflation expectations affect the future tradeoff between inflation and unemployment. A higher current inflation rate typically leads to higher inflation expectations in the future, so that it then becomes more difficult to achieve the objectives of stabilization policy.

Phelps’s contributions from the late 1960s and early 1970s radically changed our perception of the interaction between inflation and unemployment. The new theory made it possible to better understand the underlying causes of the increases in inflation and unemployment that took place
during the 1970s. A key result was that the long-run rate of unemployment cannot be influenced by monetary or fiscal policy affecting aggregate demand. Phelps’s analysis thus identified important limitations on what demand-management policy can achieve. This view has become predominant among macroeconomic researchers as well as policymakers. As a result, macroeconomic policy is carried out very differently today from what it was forty years ago.

Modern theory of capital accumulation and economic growth originated with the so-called Solow-Swan model from the mid-1950s, which shows how capital accumulation and technological change in combination lead to growth in output and consumption. But the original analysis of the model did not provide any guidance concerning appropriate rates of accumulation of physical capital and R&D. Through a number of contributions in the 1960s, Phelps focused on the intergenerational aspect of the savings problem. He showed that optimal capital accumulation, subject to a requirement of equal treatment of future generations of consumers, delivers a simple rule of thumb for long-run saving: the aggregate savings rate should equal the share of capital in national income. He explored conditions under which all generations would gain from a change in the rate of savings. He also analyzed the role of human capital for economic growth by arguing that an educated, well-trained workforce is better able to adopt available new technologies. As a result, the rate of output growth should be expected to be positively related to the average level of human capital – a proposition that has found empirical support.

**Phelps’s work on inflation and unemployment**

In the early postwar period, macroeconomics was dominated by Keynesian views on how the economy operates. According to Keynesian theory, there was no conflict between full employment and price stability. As long as the economy was below full employment, inflation
was assumed to be unaffected by an increase in aggregate demand, which could be achieved through fiscal or monetary measures. Indeed, from the Keynesian perspective the task of stabilization policy seemed almost trivial: simply keep aggregate demand high enough to avoid underemployment but not so high as to lead to excess demand for labor (overemployment) and inflation.

According to the Phillips curve (Phillips, 1958) there was instead a stable negative relationship between inflation and unemployment. This led to a revision of the standard Keynesian model of the economy. The Phillips curve implied a tradeoff between inflation and unemployment. While it was still conceivable that employment could be increased permanently using aggregate-demand policy, this would occur at the cost of higher inflation. The rise in inflation, however, would be a one-shot change from one stable level to another. The Phillips curve thus seemed to provide a menu of choice for policymakers, who could choose between inflation and unemployment according to their preferences.

There were several problems with this view. One problem was the absence of convincing microeconomic foundations for several relationships. The Phillips curve, in particular, was essentially a statistical correlation with rather tenuous theoretical underpinnings. The common interpretation, first suggested by Lipsey (1960), was that unemployment could be seen as a proxy for excess demand (supply) in the labor market, so that the Phillips curve should be regarded as an equation describing how the price (wage) level responds over time to an imbalance between demand and supply. A second problem was inherent in the view that higher employment could be permanently achieved by allowing higher inflation. This idea was in conflict with the traditional presumption in economic theory that, in the long run, real magnitudes in the economy are determined by real rather than nominal forces. A third problem was the lack of any theory
regarding the determinants of the unemployment prevailing at “full employment”, so-called frictional unemployment. Although it was generally accepted that full employment did not literally mean zero unemployment, there was no theory that specified the determinants of frictional unemployment.

Phelps’s research program from the end of the 1960s aimed at rectifying the theory of inflation and unemployment by explicitly modeling firms’ wage and price-setting behavior. Phelps brought agents’ expectations to the forefront of the analysis, made the crucial distinction between expected and unexpected inflation, and examined the macroeconomic implications of this distinction. His reformulation of the Phillips curve has become known as the expectations-augmented Phillips curve. Unlike previous studies, such as Lipsey (1960), Phelps emphasized that it was the difference between actual and expected inflation, and not inflation per se, that is related to unemployment.

Phelps’s analysis was at variance with earlier views that higher employment could permanently be achieved by inflationary demand policies. Indeed, it implied that there was no long-run tradeoff between inflation and unemployment, as there could be no permanent discrepancy between actual and expected inflation. This hypothesis of a vertical long-run Phillips curve at the equilibrium unemployment rate is one of the most influential ideas in macroeconomics over the past 50 years.¹ The hypothesis has become crucial for the conduct of monetary policy. Modern central banks generally make their decisions on interest rates on the basis of estimates of the

---

¹ The Phillips curve is typically illustrated in a figure with inflation on the vertical axis and unemployment on the horizontal axis.
equilibrium unemployment rate. In this context, Phelps also provided the first analysis of the determinants of the equilibrium unemployment rate.

Phelps’s most important contributions to the theory of inflation and unemployment consist of three papers: Phelps (1967, 1968a, and 1970b). The 1970 paper is an extension of his 1968 one and appeared in the famous monograph *Microeconomic Foundations of Employment and Inflation Theory*, usually referred to as the “Phelps volume” (1970a). In his 1967 paper, Phelps analyzed optimal demand policy when there is no long-run tradeoff between inflation and unemployment. The 1968 and 1970 papers studied wage setting and equilibrium unemployment when markets are characterized by frictions. Combined, these papers contain the core of the new insights in Phelps’s program.

**The Phillips curve and optimal inflation policy**

How should monetary (or fiscal) policy be conducted when there is a short-run – but no long-run – tradeoff between inflation and unemployment? Phelps (1967) took as given the inflation-unemployment mechanism for which he provided micro foundations in Phelps (1968a). Here, for the first time the notion of “the expected rate of inflation” was used in the Phillips-curve literature. In particular, Phelps introduced an expectations-augmented Phillips curve:

$$\pi = f(u) + \pi^e,$$

where $\pi$ is the actual rate of inflation, $\pi^e$ is the expected rate of inflation and $f(.)$ a decreasing function of unemployment, $u$. According to the equation, actual inflation depends on both

---

2 As recognized by Phelps (1968a), the idea that perfectly anticipated inflation has no real effects on the economy had appeared in earlier studies, such as Lerner (1949). However, this literature had not integrated this hypothesis within explicit models of unemployment and wage dynamics.
unemployment and the expected rate of inflation. A one percentage point increase in expected inflation leads to a one percentage point increase in actual inflation at a given rate of unemployment. The equilibrium rate of unemployment is obtained as the outcome of an expectational equilibrium with $\pi = \pi^e$, i.e., a situation with equality between the current and the expected inflation rate. That is, the equilibrium rate, $u^*$, is given by $f(u^*) = 0$. If expectations are based on observations of inflation in the recent past – so-called adaptive expectations – it follows that inflation will be increasing as long as $u < u^*$ and decreasing when $u > u^*$.\(^4\) This is often referred to as the “accelerationist hypothesis”. The term NAIRU (the non-accelerating inflation rate of unemployment) is often used as a synonym for the equilibrium unemployment rate.

Phelps’s formulation of the Phillips curve implies that the task of demand management policy is no longer to solve the static optimization problem of achieving the best attainable combination of inflation and unemployment at a given point in time. Instead, the problem is dynamic and involves finding the socially optimal paths for inflation and unemployment over time. To that end, Phelps introduced a social utility function that takes the form of a (possibly) discounted sum of instantaneous utility flows. The utility flow at each point in time depends in turn on consumption and leisure. To solve the optimization problem, policymakers have to recognize that lower unemployment and higher inflation today will raise expected inflation. This implies an intertemporal cost as it deteriorates the tradeoff between inflation and unemployment in the future. In Phelps’s model, the dynamic path of unemployment converges to a steady state where actual and expected inflation coincide and unemployment coincides with the equilibrium rate.

\(^3\) Phelps actually focused on employment rather than unemployment, but employment is of course just the mirror image of unemployment if the labor force is fixed.

\(^4\) More generally, the adaptive expectations hypothesis can be written as $d\pi^e/dt = \lambda(\pi - \pi^e)$, where $\lambda > 0$. According to the equation the expected inflation rate at each point in time is adjusted to the difference between actual and expected inflation. Combining this equation with the one in the text and holding unemployment constant gives $d\pi^e/dt = d\pi/dt = \lambda f(u)$, resulting in $d\pi/dt > 0$. 
unemployment rate. The optimization will also determine the preferred inflation rate in the steady state. The exact path depends, inter alia, on the initial expected inflation rate. If this rate is higher than the preferred steady-state level, it will be optimal to go through a period of temporary underemployment in order to bring down inflation.

This intertemporal approach to monetary policy has become standard. Contemporary academic analyses of monetary policy, as well as policy deliberations by central banks, focus on intertemporal tradeoffs, where short-run changes in activity are weighed against the effects on the possibilities to hold down both inflation and unemployment in the future. The theoretical underpinnings for the policy of inflation targeting, which many central banks have adopted since the early 1990s, are to a large extent derived from the framework developed in Phelps’s 1967 paper.5

Later work on monetary policymaking has departed from Phelps’s assumption that inflation expectations are adaptive. Based on the original work of Muth (1961) and Lucas (1972, 1973), the standard theoretical assumption now is that expectations are “rational”, i.e., they are forward looking and correct on average. Empirical studies suggest that inflation expectations do have a backward-looking component, thus pointing to the practical relevance of Phelps’s insight that both inflationary and disinflationary processes are likely to be gradual. Recent theoretical research has also attempted to explain why past inflation expectations may, after all, have a lasting effect on current inflation.6 Some of this literature has been inspired by Phelps’s own critique of the rational-expectations hypothesis. Fully rational expectations appear implausible when agents have to form expectations not only about aggregate conditions, but also about other conditions.

---

5 See Clarida, Gali and Gertler (1999) or Woodford (2003a) for surveys of current theories on monetary policy-making. Phelps (1978b) provides another account of his views on the problem of inflation planning.

6 This includes contributions such as Sargent (1999), Evans and Honkapohja (2001), Mankiw and Reis (2003), Woodford (2003b), and Orphanides and Williams (2005),
agents’ beliefs about their own beliefs and so on in an infinite chain (Phelps, 1983, Frydman and Phelps, 1983).

Microeconomic foundations for wage and price setting

Phelps (1968a, 1970b) derived aggregate wage-setting behavior from detailed modeling of the behavior of individual agents. Jobs and workers are heterogeneous and there is imperfect information on both sides of the market. Markets are assumed to be almost atomistic, but there is no “Walrasian auctioneer” that instantaneously finds the wages (and prices) that clear all markets (as went the metaphor used in earlier general equilibrium analysis). Instead, wages are set by firms that are able to exercise temporary monopsony power. Workers and firms meet randomly at a rate determined by the number of unemployed workers searching for a job and the number of vacancies, according to a function that would today be recognized as a matching function. Phelps’s work here is a precursor of the search and matching theory of unemployment, where Peter Diamond, Dale Mortensen, and Christopher Pissarides have made especially important contributions.7

In Phelps’s model, each firm anticipates that it can increase its hiring rate and decrease its quit rate by raising its relative wage, i.e., its wage relative to the average of expected wages paid by other firms. The firm tries to set a relative wage that increases with the firm’s desired net hiring rate and decreases with the current rate of unemployment. Under these assumptions, Phelps showed that there will be a unique rate of unemployment, the equilibrium rate, at which the average firm will raise its wage offer at a rate equal to the expected rate of increase in the average wage rate.

7 See, for example, Diamond (1984), Mortensen (1982a, b), Mortensen and Pissarides (1994), and Pissarides (2000).
Phelps (1968a) contains several innovations. For the first time detailed microeconomic reasoning concerning market interactions and the determination of wages and prices were introduced into mainstream macroeconomic theory. While this may seem self-evident today, it was not so at the time. On the contrary, macroeconomic relationships used in analytical models usually had the character of broad empirical generalizations, and were not based on explicit modeling of individual behavior. In Phelps’s analysis, neither demand nor supply determined the quantity of labor traded. Instead the market was typically characterized by the simultaneous existence of vacant jobs and unemployed workers, an implication of the frictions captured by the matching function.

Around the same time as Phelps presented his models of equilibrium unemployment, Milton Friedman (economics laureate in 1976) provided his well-known critique of the Phillips curve (Friedman, 1968). Like Phelps, Friedman emphasized the importance of inflation expectations for the inflation-unemployment tradeoff. Friedman shared Phelps’s view that unemployment cannot be permanently reduced by expansionary demand policy and that there exists an equilibrium rate of unemployment which is determined by the real features of the economy. In Friedman’s terminology, this rate was labeled “the natural rate of unemployment”.

Unlike Phelps, Friedman did not embed his discussion of inflation and unemployment in a formal model. It was an open question whether the expectations-augmented Phillips curve should be interpreted as a price adjustment relationship in a situation with disequilibrium between demand and supply (as was the case in Phelps’s analysis) or as an aggregate supply equation in a competitive labor market where instantaneous wage adjustments clear the labor market at all times. Under the latter interpretation, the causation runs from a deviation of actual from expected
inflation to unemployment instead of the other way around as in Phelps’s work. The idea then is that employers’ and workers’ perceptions of inflation differ. A rise in inflation induces employers to offer higher money wages. If the rise is unanticipated by workers, the higher money wage is interpreted by them as a rise in the real wage and will make them offer a larger labor supply (since labor supply is taken to depend on expected rather than actual real wages). This reduces the actual real wage (which is what matters for firms’ employment decisions) and therefore leads to an expansion of employment and output. Arguably due to its more realistic underpinnings, Phelps’s approach is the one that has had more influence on the subsequent development of models of inflation and unemployment, in particular within the so-called New Keynesian analysis of inflation and monetary policy.

Another major contribution by Phelps is his formulation of an explicit model of an imperfect labor market with frictions, job search behavior and wage-setting firms that can explain the determinants of equilibrium unemployment. Phelps (1968a) also represented the first explicit analysis of the equilibrium unemployment rate as the unemployment rate consistent with expectational equilibrium. It also provided the first formal model of so-called efficiency wages embedded in a macroeconomic framework. The general idea of efficiency wage models is that a firm may find it profitable to set its wage above the market-clearing level, thereby improving worker morale (less shirking), reducing costly labor turnover or improving the quality of the pool of job applicants. Phelps focused on the relationship between the firm’s relative wage and labor turnover, an approach further developed by others, e.g., Salop (1979), as well as by Phelps himself in his more recent work on unemployment theory; see e.g. Phelps (1994).9

---

8 This is Friedman’s own interpretation of his theory, as is clear from his Nobel lecture; see Friedman (1977).
9 Phelps often describes his analysis of wage setting in terms of “incentive wages” rather than efficiency wages. The most well-known efficiency wage model is probably that of Shapiro and Stiglitz (1984), who stressed how firms set wages to provide an incentive for employees not to shirk. See also Layard, Nickell and Jackman (1991).
In his 1968 paper, Phelps also considers the possibility that the long-run Phillips curve may be negatively sloped at very low rates of inflation, although it is vertical at higher rates. The explanation given is that, in such a situation, a reduction in the expected rate of wage increase may not translate one-to-one into a reduction in the actual rate of wage increase. The reason is that a reduction would imply cuts in the money wage level in some firms (those hit by the largest adverse demand shocks), to which there is strong psychological resistance. Phelps’s discussion here is a precursor to later work, initiated by Akerlof, Dickens and Perry (1996), who find empirical support for a negatively sloped long-run Phillips curve at rates of inflation below 1-2 per cent.

In collaboration with Sidney Winter, Phelps also made an important contribution to the theory of price setting (Phelps and Winter, 1970). The model, subsequently known as the Phelps-Winter model, seeks to explain why prices need not fully reflect short-term fluctuations in marginal costs. Although firms produce a homogeneous good in the model, they can exercise transitory monopoly power because consumers have imperfect information about the distribution of prices across sellers. A firm can therefore increase its price in the short run without immediately losing its customers, but it will have no market power in the long run (since keeping its price higher than competitors’ prices will eventually imply a loss of all customers). When setting its price, each firm trades off the gain from exploiting its short-run market power against the future reduction in profits that will occur if a price differential to competitors leads to a loss of customer stock. The optimal price path converges to a steady state where the price is higher than the marginal cost (because each firm chooses to exercise some temporary monopoly power), but lower than the static monopoly price (because this would erode the firm’s future market share). A noteworthy feature of the model is that it can generate pro-cyclical movements in the real product wage (the ratio between the wage and the product price), i.e., output may rise although
the firm’s price falls relative to the nominal wage. Nowadays, the idea that business fluctuations give rise to changes in the relationship between prices and marginal costs is commonplace in macroeconomics. The literature includes a variety of models with different mark-up predictions (see, e.g., Stiglitz, 1984, and Rotemberg and Woodford, 1999).

Subsequent research on inflation and unemployment

The consequences of imperfect information became a central theme in the “New Classical” research agenda launched by Robert Lucas in the 1970s (Lucas, 1972, 1973, 1976). In his 1972 analysis, Lucas studied how business-cycle fluctuations result from money surprises, using the island parable from Phelps (1969), but Lucas introduced rational expectations instead of the adaptive expectations used by Phelps. It turned out that Phelps’s result – that money surprises could have temporary real effects even in the absence of nominal rigidity – was robust under rational expectations. A principal conclusion was that systematic monetary policy had no role to play, as it would be built into inflation expectations and thus lead to price changes that would nullify its effects. This was forcefully argued by Sargent and Wallace (1975), who claimed that a feedback monetary policy rule could have no stabilizing effect on the economy: only errors on the part of the central bank or unexpected changes of the rule could influence output and employment.

The stabilizing power of monetary policy was the subject of an influential paper by Phelps and one of his students, John Taylor (Phelps and Taylor, 1977). Their objective was to examine whether a case could be made for stabilization policy by introducing a modest degree of wage

---

10 The explanation is that the price mark-up over marginal cost may fall in a boom. If, instead, the price is set as a fixed mark-up over marginal cost (or is equal to marginal cost as is the case with perfect competition), a rise in output is associated with a fall in the real product wage. The reason is that an increase in output (employment) implies a fall in the marginal product of labor, which tends to increase the marginal cost. With a fixed mark-up over the marginal cost, it follows that the price will then rise relative to the wage.

11 Phelps’s island parable perceives the economy as a group of islands between which information flows are costly. To learn the wage paid on a nearby island, a worker must spend the day travelling to the island to locate a wage offer rather than spending the day at work.
and price stickiness into an otherwise standard rational-expectations model. The key assumption was that monetary policy can be changed more frequently than wages and prices, so that monetary policymakers could act on the basis of a larger information set than price and wage setters. This appears to be an important real-world feature, as wage contracts are often long term. In such a framework, a negative shock to aggregate demand could generate a prolonged recession even if expectations were rational. An important result was that although systematic policy could not affect the average size of the real output gap, it could determine the variance of that gap, as well as the variance of inflation. The idea of a tradeoff between the variances of different policy objectives is of primary importance in modern discussions of monetary rules under rational expectations. The Phelps-Taylor paper is the first where such a tradeoff is derived under rational expectations.

In other contributions, Phelps studied the consequences of unsynchronized wage contracts, i.e., the empirical fact that contract decisions are typically staggered over time rather than made at the same point in time. Phelps (1970b) provided an early analysis and Phelps (1978a) showed that such staggering could substantially increase the persistence of the real effects of monetary disturbances. Taylor (1979) showed how staggered wage setting could be embedded in a model of wage and output dynamics under rational expectations. Another of Phelps’s students, Guillermo Calvo, modeled staggered prices in an economy where old prices die off asymptotically, an approach that greatly facilitated the analysis (Calvo, 1983). This particular way of modeling price adjustments has become very common in the New Keynesian literature.

The Phelps volume (1970a) was followed two years later by a new book, Inflation Policy and Unemployment Theory, dealing with the design of monetary policy (Phelps, 1972a). The

---

12 Stanley Fischer (1977) elaborated on the same basic theme, i.e., the stabilizing power of monetary policy even under rational expectations. Fischer considered overlapping labor contracts that introduced an element of wage stickiness into the model.
ambitious task was to examine the principles of optimal inflation policy and conduct a dynamic cost-benefit analysis of monetary policy. To some extent, this book served to popularize Phelps’s previous work on the expectations-augmented Phillips curve and to draw policy conclusions from the new theory. A lasting contribution from this book is the idea of hysteresis in the unemployment rate: an increase in unemployment may turn out to be (partially) irreversible, as a result of workers’ loss of skill and morale. Analogously, a temporary increase in employment may bring about lasting reductions in equilibrium unemployment as a result of favorable effects on work experience. Phelps’s idea remained largely ignored for more than a decade, but came to life again in the mid-1980s when economists struggled to understand the seemingly permanent rise in European unemployment. This research has demonstrated that unemployment persistence is a pervasive feature of many European labor markets, although the relative importance of different persistence mechanisms is still not fully understood.

The European unemployment problem also induced Phelps to dig deeper into the causes of unemployment. His analyses may be found in a number of journal articles and especially in the monograph Structural Slumps from 1994. Here Phelps turns away from the temporary expectational disequilibrium causes of unemployment and focuses on the real (structural) causes that determine equilibrium unemployment. This work forms part of a comprehensive research literature that has developed over the last twenty years, to a large extent inspired by the work of Layard, Nickell and Jackman (1991). A distinctive feature of Phelps’s later work is its emphasis on the importance of capital markets for the development of unemployment. A main hypothesis is that a higher real rate of interest tends to raise equilibrium unemployment through negative effects on the incentives of firms to invest (in physical capital, low labor turnover, customer stock, etc.). Although Phelps finds empirical support for the importance of the real rate

---

13 The adverse effects of high unemployment on human capital have been emphasized in the influential study of European unemployment by Layard, Nickell and Jackman (1991).

of interest for equilibrium unemployment, there is no general consensus in the empirical literature regarding the quantitative importance of this mechanism.

**Phelps’s work on capital accumulation**

Phelps’s perspective on the tradeoff between inflation and unemployment was fundamentally intertemporal: low current inflation can be viewed as an investment in low inflation expectations, thereby allowing a more favorable inflation-unemployment tradeoff in the future. To Phelps, the intertemporal perspective was natural, since his first research interests concerned capital accumulation. Over a ten-year period beginning in the early 1960s Phelps made a number of significant contributions in the area of capital accumulation and economic growth.

In the late 1950s and early 1960s, a view emerged in the public debate that the aggregate US savings rate was too low. A key issue in the discussion was how society should trade off the consumption of current citizens against that of future citizens. Phelps emphasized the objective of fairness among generations. Based on this view, his research examined how to best accumulate capital. He advanced the notion of *dynamic inefficiency* and used it to provide an upper bound on what the savings rate should be. Somewhat later, he returned to explicit welfare comparisons across generations of consumers and suggested a new way of thinking about saving decisions. The upshot was that savings rates could, in fact, be too low. In joint work with Robert Pollak, Phelps suggested that an induced increase in the rate of saving for all generations could be welfare-improving for all of them; see Phelps and Pollak (1968). All in all, Phelps generated a rich set of insights about optimal capital accumulation that constituted important advances in the field and have become cornerstones of capital theory.
Phelps carried out much of his analysis in the context of the growth model that is still the core of modern growth theory: the neoclassical growth model, developed in the mid-1950s by Robert Solow (1956, 1957; economics laureate in 1987) and Trevor Swan (1956). Using a stylized production structure, the Solow-Swan model describes how capital accumulation and technological progress generate output growth. In the model, the economy produces one output good, which can be used for either investment or consumption, with labor and capital as inputs. The production function has constant returns to scale but features diminishing marginal returns to the use of any given input. Investment is assumed to be a constant, exogenous fraction of output. Thus, Solow and Swan merely assumed, without a normative evaluation, how much was saved and then derived the implications for the dynamics of output. Phelps’s first question can be framed exactly in these terms: in the context of the Solow-Swan theory, what should the rate of savings ideally be? Later, Phelps analyzed broader questions, which involved a departure from the neoclassical one-dimensional setup: what form should savings take, e.g., how should investments in R&D be made, and what is the role of human capital for economic growth?

Capital theory did not start with Solow and Swan, and the analysis of optimal saving did not start with Phelps. The first fully rigorous analysis of optimal saving is the study by Frank Ramsey (1928), who formulated the individual household’s saving problem as an intertemporal maximization problem. The household chooses among different feasible consumption sequences in order to maximize a utility function that depends on a discounted stream of utility flows from consumption at different dates. A society-wide perspective on saving and capital accumulation, as well as early formal contributions to capital theory and, more generally, to competitive equilibrium theory in intertemporal models, may be found in the work of John von Neumann (1945/46), Maurice Allais (1947; economics laureate in 1988), and Edmond Malinvaud (1953).

15 In addition, Solow (1957) provided tools for an empirical evaluation of different sources of output growth, so-called growth accounting.
Later, capital theory was further developed and refined by David Cass (1965, 1972) and Tjalling Koopmans (1965; economics laureate in 1975) and others. By the late 1980s, attention turned anew to economic growth. A vibrant literature developed capital theory towards explaining the origins of technical change and the growth of knowledge, while also focusing on cross-country differences in growth. Some of Phelps’s contributions had an immediate impact on the field, but others, such as those on the roles of R&D and human capital, were far ahead of their time and had their impact only after a considerable time lag.

*Capital accumulation and intergenerational tradeoffs*

One of Phelps’s first publications is perhaps his best known contribution: his 1961 article containing the so-called golden rule of capital accumulation. Using a Solow-Swan neoclassical growth model, Phelps set out to find the most desirable long-run savings rate. His criterion was the maximization of per-capita consumption in the long run, thereby restricting attention to so-called steady states of the Solow-Swan model where the stock of capital and consumption are constant in per-capita terms. The term golden rule is a reference to the ethic of reciprocity: “Do unto others as you would have them do unto you”. The constancy of the savings ratio across generations in a steady state is thus the notion of treating others as oneself in this context. Phelps showed that the steady state with the highest consumption can be characterized by a simple prescription: the savings rate should equal the share of capital in national income. This is the golden rule, which can also be stated differently: the return to capital should equal the growth rate of output. This last expression had appeared earlier in the work of Maurice Allais on

---

17 In order to simplify the analysis, Phelps presumed that welfare depended only on consumption; for example, he abstracted from the value consumers derive from leisure.
18 As an illustration, suppose that per-capita production of output \( y \) is \( f(k) \), where \( k \) is capital per worker, that total saving is \( sy \), where \( s \) is the fraction of income saved, that capital does not depreciate, and that population grows at a constant rate \( n \). In a steady state with constant capital per worker, \( sf(k) = nk \) must hold: any investment is used to provide capital to new workers. Thus, maximization of per-capita consumption amounts to choosing \( s \) and \( k \) to maximize \( (1-s)f(k) \) subject to \( sf(k) = nk \), yielding \( f'(k) = n \). Under perfect competition, the market return to capital
intertemporal production, although Phelps’s formulation came to have a greater influence on subsequent research, since it was derived in the context of the neoclassical growth model, also allowing a formulation of the golden rule directly in terms of the savings rate.19

The golden-rule prescription specified a savings rate that was of the same order of magnitude as that in U.S. data. Therefore, Phelps’s analysis did not suggest a radical departure from prevailing macroeconomic policy. But there was also a limitation in Phelps’s (1961) analysis: it focused on the best (and equitable) long-run outcome, without exploring the costs (or, possibly, benefits) current generations would have to incur to reach this long-run state. For example, if the golden rule called for a higher savings rate and a higher long-run capital stock, then such a steady state could not be reached without the current generations having to forego consumption, which would leave them worse off than future generations. Thus, the policy path required to reach a golden-rule state may not be easy to implement politically or even desirable. Phelps, however, went on to analyze possible circumstances under which, intergenerational concerns notwithstanding, changes in the savings rate of the economy are desirable.

Phelps (1965) makes the argument that savings rates higher than the golden rule are not beneficial for any generation, independently of the ways in which consumers derive utility from consumption. The study thus examined transition paths and went beyond the earlier focus on long-run comparisons only. Here, Phelps points to the possibility of dynamic inefficiency, an important concept that others later examined in great detail and gave a complete characterization; (the price of capital services), \( r \), equals the marginal product of capital: \( r = f'(k) \). Thus, optimal saving requires \( r = n \). Since \( sf(k) = nk = rk \), we deduce that capital income, \( rk \), must equal \( sy \), and hence that \( s = rk/y \). The results generalize easily to allow depreciation of capital at rate \( \delta \) and “labor-augmenting” technical progress at rate \( g \). In this case we still obtain \( sy = rk \), where now \( r = n + g + \delta \). In other words, the net real interest rate, \( r - \delta \), should equal the growth rate of total output, \( g + n \).

19 See Allais (1947) and also Allais (1962), which discusses his earlier work. At about the same time as Phelps (1961), golden-rule results were also published by Desrousseaux (1961), Robinson (1962), von Weizsäcker (1962), and Swan (1964). The latter paper, presented at a conference in 1960, derives the golden rule within the Solow-Swan growth model and expresses it, as Phelps does, in terms of the savings rate. The problem of optimal capital accumulation is also analyzed in Malinvaud (1953).
see, in particular, Cass (1972). Dynamic inefficiency has become a standard concept in contemporary normative theory of economic growth. Formally, a capital path is dynamically inefficient if the path of saving can be changed so as to strictly increase consumption at some point in time without lowering it at any other point in time. Thus, dynamic inefficiency implies overaccumulation of capital and, when this occurs, it constitutes a drastic market failure in the sense that resources are simply “left on the table”. In his 1965 paper, Phelps showed in particular that whenever the long-run savings rate is above the golden-rule rate, there is dynamic inefficiency. In his paper, Phelps acknowledged that Koopmans had helped with the proof; the concept is often referred to as “Phelps-Koopmans dynamic inefficiency”. The intuition is straightforward. A lowering of the savings rate increases consumption in the short run, since it will take time for the capital stock, and hence production, to adjust downward significantly. But if the initial capital stock is above the golden-rule level, then the drop in savings will also increase consumption in the long run, since the golden rule, by definition, maximizes long-run consumption: despite a lower long-run capital stock, consumption will increase. The striking conclusion, thus, is that regardless of which intergenerational utility perspective is adopted, a savings rate above the golden-rule rate is too high.

If the savings rate lies below the golden-rule rate, an argument for higher saving has to rely on a more careful consideration of the welfare outcomes for different generations. In joint work with

---

20 The argument can be made precise by the following example, where, in addition to the symbols used in footnote 18, \( A \) is the level of technology. Consider a discrete-time economy with neither population nor technology growth, where \( f(k) = Ak^\alpha \). Thus, the resource constraint is \( c(t) + k(t+1) = Ak(t)^\alpha + (1-\delta)k(t) \). Suppose that the economy is in a steady state with a constant level of capital, \( k^* \), so that consumption is \( c^* = A(k^*)^\alpha - \delta k^* \). Now consider decumulating capital from \( k^* \) to \( k^* - e \) in one period and keeping it there. Consumption in all future periods will then be \( A(k^* - e)^\alpha - \delta (k^* - e) \). Differentiating with respect to \( e \) and evaluating at \( e = 0 \), one deduces that consumption must increase if \( \alpha A(k^*)^{\alpha - 1} < \delta \) and \( e \) is small enough. Moreover, initial consumption rises to \( c^* + e \).

21 Phelps was not the first to discuss efficiency of capital accumulation paths. Based on a given set of intertemporal prices, Malinvaud (1953) and later others provided conditions for efficient aggregate savings. However, these studies did not provide guidelines as to which prices should be used. Phelps’s characterization, in contrast, is made directly in terms of the given capital accumulation path.
Robert Pollak this is the route Phelps took next; see Phelps and Pollak (1968). The underlying idea is that each consumer cares not only about his own consumption level, but also about the consumption levels of his offspring: there is altruism, so that individuals also save to give bequests. Based on a specific view regarding consumers’ valuation of their own consumption and that of future generations, Phelps and Pollak showed that equilibrium savings rates can sometimes be too low.

Phelps and Pollak assumed that each generation values their offsprings’ consumption, but not in the same way as the offsprings do. Suppose that you care about your child’s and grandchild’s consumption levels equally, but that you care more about your own, and that your child feels the same way about his consumption relative to that of his child and grandchild. Then, Phelps and Pollak argued, you would disagree with your child’s savings choice: you would want your child to save more than he would choose to do. Formally, these differential valuations of different generations’ consumption utilities were cast in terms of discounting weights; Phelps-Pollak discounting has more recently been dubbed quasi-hyperbolic, or quasi-geometric, discounting. Consequently, it is natural to view the savings decisions of different generations of consumers as determined in a non-cooperative game between generations. Phelps and Pollak carried out their analysis within such a framework and showed that, absent government intervention, bad outcomes could arise. In particular, a policy-induced increase in all generations’ savings rates could increase the welfare of all generations. Even though a forced rise in your savings rate is, by itself, not in your interest, if it is accompanied by a forced rise in your offsprings’ savings rates,

---

22 Time-inconsistent preferences had been analyzed earlier. Ramsey (1928) considers the possibility, whereas Strotz (1956) makes the first formal analysis. Pollak (1968) comments on some aspects of Strotz’s analysis. Quasi-hyperbolic discounting, when time is discrete, is characterized by two different discounting parameters used by a decision-maker at any time \( t \): a (short-run) rate that applies between \( t \) and \( t+1 \), and a (long-run) rate that applies between any two adjacent later dates. Thus, when these two parameters differ, there is disagreement. A decision-maker at time \( t \) will attach a different relative weight to utility flows in \( t+s+1 \) relative to those in \( t+s \) than will a decision-maker at \( t+s \), for any positive \( t \) and \( s \). A “present bias” obtains when the short-run discount rate is higher than the long-run discount rate.
there is a counteracting effect, which can dominate. This provides a “paternalistic” argument in favor of forced savings, suggesting that governments should intervene in private consumption decisions. This argument has been used to justify, or explain, an active role for governments in old-age savings plans (pension systems). Recently, the Phelps-Pollak model has also been recast in terms of the consumption choice of a single individual with “time-inconsistent” preferences (say, the consumer who wants to stop smoking, only not right now). As such, the model has had a far-reaching impact; see, e.g., Laibson (1997) and the recent psychology-and-economics literature.

Phelps’s work on optimal savings was followed by important later contributions, by others as well as by Phelps himself. Cass (1965) and Koopmans (1965) considered the neoclassical growth model using dynastic settings, with preferences of the form used in Ramsey (1928), and they fully characterized optimal savings paths. These studies led to “modified golden rules”, where utility discounting was incorporated into the rule. Another set of studies followed Peter Diamond’s (1965) influential analysis of an overlapping-generations economy with a neoclassical production structure, where Phelps’s setting was augmented with a fully specified population structure. Here, each generation is endowed with a utility function over consumption streams but has no altruistic motives; consumers then maximize utility using perfectly competitive loan markets. Diamond showed that competitive equilibria could indeed deliver dynamic inefficiency. Phelps and Riley (1978) also considered an overlapping-generation structure and examined capital accumulation paths that are not restricted to constant savings rates. The intergenerational equity consideration is formalized by adopting a Rawlsian social

---

23 It is, for example, straightforward to show that under logarithmic utility and a savings technology that is linear in the amount saved, equilibrium savings in the absence of government intervention are too low.

24 Abstracting from technology and population growth, a steady state resulting from intertemporal utility maximization is characterized by a savings rate that is constant and that modifies Phelps’s golden rule \( r = \delta \) to \( r = \delta + \rho \), where \( \rho \) is the rate at which dynasties discount their utility flows over time. Thus, given that \( r=f(k) \) is decreasing in \( k \), less capital accumulation is called for due to discounting.
welfare function, where the objective is to maximize the well-being of the generation which is least well off.\textsuperscript{25} The study argues that golden-rule outcomes are not generally optimal but that a path with net capital accumulation actually can be, even though any capital accumulation moves resources away from the initial generation.

Phelps also initiated the analysis of capital accumulation under uncertainty (Phelps, 1962). Uncertainty, especially about the return from capital accumulation, is an important real-world feature. Phelps examined a Ramsey-style model with an individual consumer who faces stochastic returns to saving. Phelps used his model to gain insights into why savings rates seem to differ among consumers with different sources of income.

\textit{Accumulation in the form of R&D and human capital}

In another early contribution, Phelps examined endogenous research and development (Phelps, 1966b). Again, the perspective was not on positive analysis but on determining optimal research investment rates. Phelps examined a neoclassical model with capital, labor, and the level of technology as variable production factors. Within this framework, Phelps again derived, as a special case, a golden rule of accumulation of physical capital for maximal steady-state consumption, which is unaffected by the endogeneity of technological progress. He also derived a golden rule of investment in research for maximal steady-state consumption. The rule prescribes the equalization of the return to investments in technology to the rate of growth. Phelps concluded that the golden rule could be generalized: the rates of return of all kinds of investments should be equal to the output growth rate.\textsuperscript{26}

\textsuperscript{25} John Rawls’s book \textit{A Theory of Justice} (1971) proposes a “maximin” criterion according to which the welfare of the least well off individual in society should be maximized.

\textsuperscript{26} Inspired by the earlier work of Charles Kennedy (1964) - recently revisited in a number of studies by Daron Acemoglu (see, e.g., Acemoglu, 2002) – Drandakis and Phelps (1966) made another contribution to the analysis of optimal research efforts by studying the choice of the “direction” of research: whether it should be labor-saving or capital-saving.
The endogenous-growth literature that later emerged differed from Phelps’s work primarily in that it had a positive angle: it developed decentralized models of endogenous technological change aimed at understanding growth in a broad cross section of countries.\textsuperscript{27} Parts of this literature (in particular, Aghion and Howitt, 1999) build directly on a theoretical model of human capital and the diffusion of technology by Richard Nelson and Phelps (1966). The Nelson-Phelps analysis focuses on the stock – as opposed to the accumulation – of human capital as a key factor behind technology growth in the sense that an educated and knowledgeable workforce is better able to adopt available new technology. In the empirical growth literature, the Nelson-Phelps setting has provided a means of formalizing technological catch-up across countries, whereby technologically less advanced countries adopt the technologies of the more advanced ones, and do so more efficiently, the more educated the workforce in the adopting country.\textsuperscript{28} The analysis thus explains why data indicate that output growth is more related to the stock of human capital than to its rate of growth. The model also helps explain why skill premia – the higher wage rates enjoyed by skilled, or educated, workers – tend to be higher in times of rapid technological change: an educated workforce is able to assimilate technological advances more rapidly. Such reasoning has been used to interpret the recent increase in returns to education that has taken place in many countries, in particular the US.\textsuperscript{29} To the extent that there are important spillovers (“externalities”) in the adoption of technologies, the returns to education might not be fully reflected in the skill premium. Thus, Nelson and Phelps argued, their model suggests a possible reason for subsidizing education.

\textsuperscript{27} Another difference is that most of the endogenous-growth literature considers cases where the long-run rate of technology growth is endogenous. In most of Phelps’s work, in contrast, since researcher-workers are assumed to be essential for generating technical change, the rate of population growth, which is exogenous, ultimately pins down the long-run growth rate of technology.

\textsuperscript{28} See, e.g., Benhabib and Spiegel (1994) and Barro and Sala-i-Martin (2004).

\textsuperscript{29} For a discussion of these issues, see, e.g., Hornstein et al. (2005).
Phelps (1966b) assumes that the long-run rate of growth of technology is ultimately pinned down by the growth rate of the population. Thus, the maintained view, elaborated on in Phelps (1968b), is that the accumulation of “ideas” fundamentally benefits from population growth. The logic is simple: people are required in order for new ideas to be developed, but once a productive and valuable idea is born, it can be transferred to others essentially at no cost. Thus, the amount of new ideas per capita will grow with population size. Recent contributions to growth theory by Charles Jones (see, e.g., Jones, 1995) revisit Phelps’s arguments that population growth is central for long-run technology growth.

Phelps’s other contributions

Although Edmund Phelps is best known for his work in macroeconomics, his contributions to labor economics and public finance also deserve mention. Among other contributions he initiated the literature on statistical discrimination, derived new results regarding the structure of optimal income taxation, and examined the properties of an optimal inflation tax.

Phelps’s ideas concerning statistical discrimination were outlined in his monograph (1972a) and formalized in an article (1972b). Around the same time, Kenneth Arrow published equally influential papers on statistical discrimination (Arrow, 1972a, 1972b, 1973). These studies by Phelps and Arrow – both commonly referred to as seminal contributions to the theory of statistical discrimination – emphasize that unequal treatment of equally productive workers can arise when employers have imperfect information about individual worker characteristics. When individual productivity is measured with error, it may be worthwhile to use group data – information on average productivities in the group to which an individual belongs – so as to

---

30 Phelps explains the basic intuition with the following example (Phelps 1968b, pp. 511-512):
“One can hardly imagine, I think, how poor we would be today were it not for the rapid population growth of the past to which we owe the enormous number of technological advances enjoyed today... If I could re-do the history of the world, halving population size each year from the beginning of time on some random basis, I would not do it for the fear of losing Mozart in the process.”
improve predictions about an individual worker’s productivity. A consequence of such behavior is that individuals with the same characteristics may be treated differently.

Phelps’s most noteworthy contribution to public finance deals with optimal taxation of wage income (Phelps, 1973a). Following Mirrlees (1971), Phelps assumed that whereas workers have different wage rates per hour, the policymaker can only observe wage incomes, not wage rates (or hours worked). A striking result of this analysis is that the marginal tax rate should approach zero at the very top of the income distribution (even though the average tax rate may well be high at the top). A similar finding was later obtained by Sadka (1976) and it is often referred to as the “Phelps-Sadka result”.

The idea that inflation can be viewed as a tax on the holders of nominal assets has a long history in economic doctrine. Phelps (1973b) notes that, from a public finance perspective, inflation is a source of tax revenue for the government, and hence it should be chosen optimally in conjunction with other taxes. He thus argued in favor of a positive (but modest) inflation rate in order to balance the distortions from different taxes against each other.

Summary

Edmund Phelps’s contributions have had a decisive impact on macroeconomic research. They have also fundamentally influenced the practice of macroeconomic policymaking. The focus on the intertemporal nature of macroeconomic policy tradeoffs provides a unifying theme for his work on inflation and unemployment as well as on capital accumulation.

Toward the end of the 1960s, Phelps launched a research program with the objective of rebuilding Keynesian macroeconomic theory based on explicit modeling of the behavior of
individual firms and households in a world of imperfect information and market frictions. Phelps developed the first models of the expectations-augmented Phillips curve and equilibrium unemployment. These models fundamentally changed the analysis of the relationship between inflation and unemployment. They also provided a completely new setting for the analysis of monetary (and fiscal) demand management policy. The resulting insights changed the views, among researchers as well as policymakers, as to what such policies could achieve and how they should achieve it.

Phelps contributions to the analysis of optimal capital accumulation demonstrated conditions under which all generations would benefit from changes in the aggregate savings rate. He also pioneered research on the role of human capital for technology and output growth. His composite contributions in this area have had a profound impact on subsequent research and have improved our understanding of the process of economic growth.
References


