W

Wade, John (1788–1875)

Daniele Besomi

Abstract

John Wade was born in London to working class parents. He worked for more than a decade as a journeyman wool sorter, then he 'wrote his way out of obscurity' (Harling, P. 2004. Wade, John (1788–1875). Oxford dictionary of national biography. Oxford: Oxford University Press). He was encouraged by Francis Place to engage in journalism: his first venture was a penny newspaper, *The Gorgon,* published in 1818–19 on money lent by Bentham and Henry Bickersteth (later Baron Langdale). Wade's articles are reputed to be well informed and detailed, so that *The Gorgon*'s influence surpassed its limited circulation. It attempted to find a junction point between, on the one hand, radical reformers and trade unionists, to which group Wade belonged, and, on the other hand, moderate reformers, with particular reference to the possible use of utilitarian doctrines to improve the condition of the labouring classes.

Keywords

Bentham, J.; Capital–labour relations; Cobweb theorem; Place, F.; Speculation; Technological progress; Utilitarianism; Wade, J.

JEL Classifications

B31

In 1819 Wade published anonymously in two-penny sheets *The Black Book*, where he documented in detail the revenues and privileges of the 'parasitic' classes: clergy, aristocracy, and anyone connected with the government. The book (1820), later qualified as a handbook of radical agitators, was very successful, with over 50,000 copies sold in its various editions. In 1828 Wade joined the staff of the newly founded *Spectator*, and in the course of his life he wrote several inquiries and a number of compendia on such topics as *Manchester Massacre* (1819), *A Political Dictionary . . . Chiefly Designed for the Use of Members Of Parliament. . .* (1821), *A Popular Digest of the Laws of England* (1826, 24th edition in 1869), *Digest of Facts and Principles on Banking* (1826), *British History, Chronologically Arranged* (1839), *Principles of Money, with their Application to the Reform of the Currency and of Banking* (1842), the last book on *Harlotry and Concubinage* being dated 1859. His juvenile positions became more widely accepted in later years, and in 1862 he was granted a £50 annual pension by Lord Palmerston. He died in Chelsea in 1875.

Wade's *History of the Middle and Working Classes* (1833) offers interesting insights into the nature of capital and its relation to labour, but especially on periodic fluctuations in prices and activity. While most of Wade's contemporaries focused on crises, conceived either as accidental events or, more

Macmillan Publishers Ltd (ed.), *The New Palgrave Dictionary of Economics*,
https://doi.org/10.1057/978-1-349-95189-5

rarely, as periodically returning, Wade offered one of the earliest, if not *the* earliest, dynamic models of endogenous cycles in individual industries. Having recognized the existence of a 'commercial cycle' showing some 'periodic regularity' and recurring every five to seven years (Wade 1833, pp. 211, 255), he offered an explanation consisting essentially in a cobweb-like mechanism. Price movements trigger changes in both demand and production, which in turn react against the original movement. When prices rise, demand falls while production increases. Supply thus outpaces demand, and prices fall, setting off the opposite movement (1833, p. 254). The cycle results from the system's tendency to correct price fluctuations. The assumption, implicit in this part of the argument, that reactions are either slow or lagged was made explicit a few pages later where Wade observed that the introduction of new machinery takes some time to fully develop its consequences on production and prices (1833, p. 257).

On top of this mechanism, Wade considered a number of relieving or aggravating circumstances, including foreign competition, changes in fashion or technological progress. Notably, he saw 'illusive speculation' of an 'epidemic character', addressed 'to the passion and not to the reason of mankind', as the generator of crises. This argument was common at the time, but while his contemporaries saw in this the root of the problem, Wade was adamant that this was an aggravating cause only, superimposing onto the main mechanism.

See Also

- ▶ Bentham, Jeremy (1748–1832)
- ▶ Cobweb Theorem
- ▶ Utilitarianism and Economic Theory

Selected Works

1820. *The black book, or, corruption unmasked!*. . .. London: Fairburn.

1833. *History of the middle and working classes.* London: Effingham Wilson. Reprinted: New York: Kelly, 1966.

Bibliography

Besomi, D. 2008. John Wade's early endogenous dynamic model: 'Commercial cycle' and theories of crises. *European Journal of the History of Economic Thought* 15(4): 611–639.

Edmund, J.C., and R. Frow. 1994. John Wade and the black book. *Labour History Review* 59(2): 55.

Harling, P. 2004. Wade, John (1788–1875). In *Oxford dictionary of national biography*. Oxford: Oxford University Press.

von Bergmann, E. 1895. *Die Wirtschaftskrisen: Geschichte der nationalökonomischen Krisentheorien*. Stuttgart: W. Kohlhammer.

Zegger, R.E. 1979. Wade, John. In *Biographical dictionary of Modern British radicals*, vol. 1, ed. J.O. Baylen and N.J. Gossman. Hassocks: Harvester Press.

Wage Curve

David G. Blanchflower and Andrew J. Oswald

Abstract

This article summarizes evidence for the existence of a wage curve – a downward-sloping relationship between the level of pay and the local unemployment rate – in modern micro data. At the time of writing, the curve has been found in 40 nations. Its elasticity is approximately −0.1.

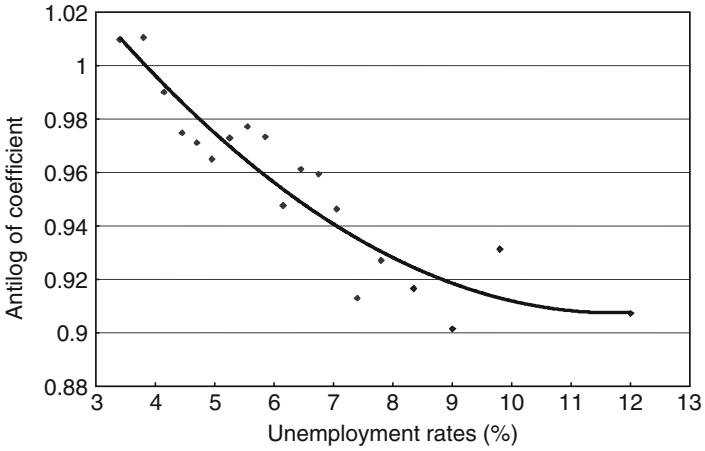
Keywords

Labour economics; Microfoundations; Phillips curve; Phillips, A. W.; Power laws; Unemployment; Wage curve

JEL Classifications

J31

The wage curve is a statistical regularity or empirical 'law' of economics. It traces out, as in Fig. 1, a downward-sloping relationship between wages and local unemployment. Its elasticity is approximately −0.1. Although this kind of downward-sloping shape has since been replicated in many other nations and by many other authors, Fig. 1



Wage Curve, Fig. 1 United States wage curve, 1963–1987 (*Source*: Blanchflower and Oswald (1994, p. 134))

here is reproduced from work on US data by Blanchflower and Oswald (1994, p. 134). The *y* axis, here labelled 'antilog', is a measure of the level of pay.

As an example, consider two regions within a country. Assume Region A's unemployment rate is double that in Region B. The wage-curve finding states that a worker's wage will then be 10 per cent lower in Region A than the wage of an identical worker in Region B.

To understand the wage curve's place in intellectual history, it is useful to go back to one of the oldest questions in economics, namely, that of how the price of labour is affected by the unemployment rate. Following an empirical tradition begun by the New Zealand economist A.W. Phillips (1958), this issue has traditionally been studied with aggregate time-series methods. Although its robustness is still questioned, the Phillips curve, which is a relationship between wage growth and unemployment, has become part of the bedrock of macroeconomics textbooks. Sargan (1964) pointed out that it was possible to view the Phillips curve as being consistent with a steady-state solution where the level of pay depends on the level of unemployment.

Blanchflower and Oswald (for example, 1994, 1995) were among the first to argue instead for the use of microeconomic data in such a setting. Their book does not study the Phillips curve, nor does it use aggregate data. Instead, using samples of individual workers, the authors document the existence of a logarithmic curve – what physicists would call a power law – linking the level of the wage to the unemployment rate in the local area. They conclude that in 16 nations, including the United States, the data are well described by a wage curve with an unemployment elasticity of approximately −0.1.

Since then, those conclusions have been checked, and largely replicated, by other researchers and on different nations' data. Examples include Hoddinott (1996) for the Côte d'Ivoire; Janssens and Konings (1998) for Belgium; Sabin (1999) for China; Bellmann and Blien (2001) for Germany; and Garcia-Mainar and Montuenga-Gomez (2003) for Spain. A recent study by Sanz-de-Galdeano and Turunen (2006) has used a large longitudinal data-set on workers across the Eurozone and, once again, obtained a similar elasticity.

Evidence for a wage-curve pattern has been found in more than 40 countries. Its existence in the United States, however, is currently viewed as somewhat more controversial. One reason is that Blanchard and Katz (1997) argue for a Phillips curve, rather than a wage curve, in United States data. Staiger et al. (2002) and Card and Hyslop (1997) also report a high level of autoregression in US wages. In contrast, Hines et al. (2001) conclude that a wage curve specification has a more natural theoretical

interpretation and fits the data (hours as well as wages) better than the Phillips curve specification. The authors produce evidence of wage curves using annual and hourly earnings from the 1977–2000 March Current Population Survey files. The authors also uncover wage curves in the Panel Study of Income Dynamics (PSID). Using the PSID, Hines, Hoynes and Krueger suggest that a three percentage point decline in the unemployment rate is associated with a 4 per cent increase in real wages, which translates into an elasticity similar to the Blanchflower–Oswald number. Recently, Blanchflower and Oswald (2005) returned to the topic of the wage curve, and, in modern US data, argued that the United States has a long-run wage curve with the usual elasticity of −0.1 but that their 1994 book should have paid more attention to the high degree of autoregression in US state wages.

The wage curve seems relevant beyond its implications for labour economics. First, macroeconomic analysis has for some decades stressed the need for microeconomic foundations. Second, some macroeconomics textbooks make extensive theoretical use of a wage curve (at the aggregate level), but do not provide evidence for it.

Wage curves have been reported for Argentina, Australia, Austria, Belarus, Belgium, Brazil, Bulgaria, Burkina Faso, Canada, Chile, China, Côte d'Ivoire, Czech Republic, Denmark, East Germany, Estonia, Finland, France, Great Britain/UK, Holland, Hungary, India, Ireland, Italy, Japan, Latvia, New Zealand, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Turkey, USA and West Germany. These studies are summarized in Blanchflower and Oswald (2005). A meta-analysis – on a sample of 208 wage–unemployment wage curve elasticities from the literature – by Nijkamp and Poot (2005, p. 445) concludes that

> the wage curve is a robust empirical phenomenon … but there is … evidence of publication bias. There is indeed an uncorrected mean estimate of about −0.1 for the elasticity. After controlling for publication bias by means of two different methods, we estimate that the 'true' wage curve elasticity at the means of study characteristics is about −0.07.

Why the wage curve exists, however, is not so well understood. One way to rationalize such a pattern is to appeal to non-competitive theories of the labour market – for example, to the idea of a no-shirking condition or a Nash bargaining-power locus. According to this kind of analytical framework, high local unemployment makes life tougher for workers (because, for example, they will find it harder to obtain work if laid off by their current employer), and therefore it is not necessary for employers to remunerate them so generously. The wage curve is then potentially an important element of a theory of equilibrium in the labour market such as in Shapiro and Stiglitz (1984) or Pissarides (2000).

Whatever the correct theoretical interpretation, new empirical results continue to emerge. Even in South Africa, where unemployment rates have run as high as 30 per cent, Kingdon and Knight (2006) conclude that there is a wage curve with an elasticity of −0.1. Although its conceptual foundations will go on being debated, and more research, especially for the United States, is required, the wage curve appears to be a pattern that holds in many nations.

See Also

- ▶ Efficiency Wages
- ▶ Phillips Curve (New Views)

Acknowledgment We thank Bruce Weinberg for helpful suggestions.

Bibliography

Bellmann, L., and U. Blien. 2001. Wage curve analyses of establishment data from Western Germany. *Industrial and Labor Relations Review* 54: 851–863.

Blanchard, O., and L.F. Katz. 1997. What we know and do not know about the natural rate of unemployment. *Journal of Economic Perspectives* 11(1): 51–72.

Blanchflower, D.G., and A.J. Oswald. 1994. *The wage curve*. Cambridge, MA: MIT Press.

Blanchflower, D.G., and A.J. Oswald. 1995. An introduction to the wage curve. *Journal of Economic Perspectives* 9(3): 153–167.

Blanchflower, D.G., and A.J. Oswald. 2005. The wage curve reloaded. Working Paper No. 11338. Cambridge, MA: NBER.

Card, D., and D. Hyslop. 1997. Does inflation 'grease the wheels' of the labor market? In *Reducing inflation: Motivation and strategy*, NBER Studies in Business Cycles, ed. C.D. Romer and D.H. Romer, Vol. 30. Chicago/London: University of Chicago Press.
Garcia-Mainar, I., and V. Montuenga-Gomez. 2003. The Spanish wage curve: 1994–1996. *Regional Studies* 37: 929–945.
Hines, J., H. Hoynes, and A.B. Krueger. 2001. Another look at whether a rising tide lifts all boats. In *The roaring nineties: Can full employment be sustained?* ed. A.B. Krueger and R.M. Solow. New York: Russell Sage Foundation.
Hoddinott, J. 1996. Wages and unemployment in an urban African labour market. *Economic Journal* 106: 1610–1626.
Janssens, S., and J. Konings. 1998. One more wage curve: The case of Belgium. *Economics Letters* 60: 223–227.
Kingdon, G.G., and J. Knight. 2006. How flexible are wages in response to local unemployment in South Africa? *Industrial and Labor Relations Review* 59: 471–495.
Nijkamp, P., and J. Poot. 2005. The last word on the wage curve? A meta-analytic assessment. *Journal of Economic Surveys* 19: 421–450.
Phillips, A.W. 1958. The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica* 25: 283–299.
Pissarides, C.A. 2000. *Equilibrium unemployment theory.* 2nd ed. Cambridge, MA: MIT Press.
Sabin, L. 1999. The development of urban labour markets: China's urban wage curve, 1980–92. *Journal of Development Studies* 35(3): 134–152.
Sanz-de-Galdeano, A., and J. Turunen. 2006. The euro area wage curve. *Economics Letters* 92: 93–98.
Sargan, J.D. 1964. Wages and prices in the United Kingdom: A study in econometric methodology. In *Econometrics and quantitative economics*, ed. D.F. Hendry and K.F. Wallis. New York: Basil Blackwell, 1984.
Shapiro, C., and J. Stiglitz. 1984. Equilibrium unemployment as a worker discipline device. *American Economic Review* 74: 433–444.
Staiger, D., J.H. Stock, and M.W. Watson. 2002. Prices, wages and the U.S. NAIRU in the 1990s. In *The roaring nineties: Can full employment be sustained?* ed. A. Krueger and R. Solow. New York: Russell Sage Foundation.

Wage Flexibility

Richard Jackman

The importance of wage flexibility arises from the fact that, in a wide range of economic models, there is an inverse relationship between wages and employment. Unemployment is thus associated with wages in excess of the full employment level, and the persistence of unemployment then depends on how quickly wages adjust in the face of unemployment. It is often argued that if wages were very flexible, unemployment would be eliminated quickly and automatically by wage cuts, and that consequently any persistence of unemployment must be ascribed to wage inflexibility.

While wage inflexibility plays a crucial role in explaining unemployment in both Classical and Keynesian models, the mechanism through which it does so is quite different in the two cases. Following Barro and Grossman (1971) and Malinvaud (1977) it is useful to distinguish 'Classical' from 'Keynesian' unemployment. Classical unemployment occurs where the real wage exceeds the marginal product of labour at full employment, so that it is not profitable for firms to employ the whole labour force. It can only be reduced by cuts in real wages which make it profitable for firms to take on more workers at the margin.

Keynesian unemployment is caused by a deficiency of aggregate demand, but in most standard presentations of the Keynesian model aggregate demand is determined, to a greater or lesser extent, in nominal terms so that a cut in money wages, and hence in prices, tends to raise real aggregate demand. Thus it is the inflexibility, or downward rigidity, of money wages which is the crucial assumption in explaining the persistence of unemployment in standard presentations of the Keynesian system. (For a very full documentation of this point see Leijonhufvud 1968.)

Wage bargaining is generally conducted in money terms, and wage flexibility is thus generally interpreted in terms of the responsiveness of money wage settlements to changes in economic conditions. But the effectiveness of money wage flexibility in reducing unemployment depends on the interaction of wage-setting and price-setting behaviour. As Keynes stressed in the *General Theory* (1936, chs 2 and 19), if a change in money wages leads to an equi-proportionate change in prices, as the standard economic theory of competitive markets might lead one to expect, it

will leave the real wage unchanged. Thus, in the Keynesian system, the wage bargain has no direct effect on the real wage. At the other extreme, in their general disequilibrium model, Barro and Grossman (1971) take the price level as fixed. In their model a fall in money wages will reduce real wages but, because there is no fall in prices, there is no stimulus to aggregate demand, and hence a fall in money wages will not help remove Keynesian unemployment.

Price-setting behaviour is important for a second reason. While wage bargains are generally conducted in money terms, it is now generally accepted that what is at issue is the real wage. There is much empirical support for the theoretical proposition that workers do not suffer from 'money illusion' (especially in countries which have had some experience of inflation), and the money wage claim is best regarded in terms of some desired real wage to be attained in the wage bargain.

The desired outcome of the wage bargain may thus be written

$$w^* = p + q - \alpha_1(u - \overline{u}), \alpha_1 > 0 \qquad (1)$$

where all variables are measured in logarithms, w^* is the desired money wage, p the price level, q labour productivity, u the unemployment rate and $\bar{u}$ a measure of 'equilibrium' unemployment in a sense to be defined below.

Equation (1) is sufficiently general to be consistent with a number of models of wage determination. Under perfect competition, it describes the equilibrium wage, given the size of the labour force, in which case $\bar{u}$ represents frictional and voluntary unemployment, determined by search behaviour, work-leisure preferences and the like. In models in which wages are not necessarily set to clear the market, the impact of trade union bargaining power or other noncompetitive influences which shift the wage equation can be captured in $\bar{u}$.

In general, wages do not adjust instantaneously to the desired level, in part because perceptions, or expectations, of the relevant variables may be slow to adjust (Friedman 1968) and in part because of rigidities in the adjustment process itself, associated for example with the existence of wage contracts (Fischer 1977; Taylor 1980). In a simplified representation, actual wages might be determined according to

$$w = \beta_1 w^* + (1 - \beta_1) w_{-1}, \quad 0 < \beta_1 < 1 \qquad (1')$$

where w is the actual, and w_{-1} the one period lagged, money wage.

The price equation may be written

$$p^* = w - q - \alpha_2(u - \overline{u}), \alpha_2 > 0 \qquad (2)$$

where p^* is the firm's desired price, $(w - q)$ is a measure of unit cost and α_2 measures the impact of the level of economic activity on the price mark-up. (The constant term in the equation is suppressed, but changes in, e.g., material prices can be represented by a change in q.) Equation (2) is consistent with price-setting behaviour by firms operating in competitive or noncompetitive markets (with a given degree of monopoly power).

Product prices may not adjust instantaneously due to slow adjustment of perceptions (or expectations), transactions costs

$$p = \beta_2 p^* + (1 - \beta_2) p_{-1}, 0 < \beta_2 < 1. \qquad (2')$$

These equations define the adjustment behaviour of wages and prices

$$\Delta w = \frac{\beta_1}{1 - \beta_1}[p + q - w - \alpha_1(u - \overline{u})] \qquad (3)$$

$$\Delta p = \frac{\beta_2}{1 - \beta_2}[w - q - p - \alpha_2(u - \overline{u})]. \qquad (3')$$

For equilibrium ($\Delta w = \Delta p = 0$) we evidently require

$$\left.\begin{array}{l} w = p = q \\ u = \quad \overline{u} \end{array}\right\} \qquad (4)$$

with unemployment at the equilibrium rate and real wages equal to labour productivity.

To examine the response of the system to a change in aggregate demand, we assume for simplicity that nominal aggregate demand (m) is determined exogenously and that unemployment responds to real aggregate demand according to

$$u = \overline{u} - \frac{1}{\gamma}(m - p), \gamma > 0 \tag{5}$$

Substituting (5) into (3) and (3′) allows the wage-price system to be converted to a representation of the economy in terms of money wages and unemployment

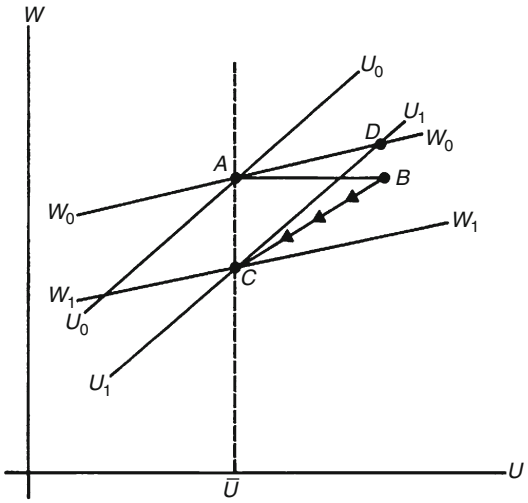
$$\Delta w = \frac{\beta_1}{1 - \beta_1}[m + q - w + (\gamma - \alpha_1)(u - \overline{u})] \tag{6}$$

$$\Delta u = \frac{\beta_2}{\gamma(1 - \beta_2)}[w - q - m - (\gamma + \alpha_2)(u - \overline{u})] + \Delta\overline{u} - \frac{1}{\gamma}\Delta m \tag{6'}$$

new equilibrium loci w_1w_1 and u_1u_1 relating to the reduced level of nominal demand, m_1.

The equilibrium of this system is given by equation (4), as before, together with $m = p$. Its dynamic behaviour is depicted in Fig. 1. The equilibrium loci $\Delta w = 0$ and $\Delta u = 0$ are depicted, at some given level of demand m_0, by the lines w_0w_0 and u_0u_0 respectively with the equilibrium of the system at point A. (The ww locus is drawn upward sloping since empirically one would expect γ to be greater than α_1.) If demand is now reduced to some lower level (m_1) initially, with given wages and prices, unemployment will rise and the system will move to point B. The higher unemployment will cause wages and prices to fall and the economy will move along the path BC, the final equilibrium position C being defined by the intersection of the

The crucial issue is the speed at which the economy progresses along the path BC. This speed is jointly determined by the parameters of equations (6) and (6′) and hence on the flexibility of prices (β_2) as much as of wages (β_1). The algebraic solution to equations (6) and (6′) is standard, and while there is no simple analytical expression for the speed of adjustment it can be confirmed that adjustment is quicker the larger the values of the demand effects on wages and prices (α_1 and α_2) and the greater the flexibility of wage and price adjustment (β_1 and β_2).



Wage Flexibility, Fig. 1 Wage and unemployment dynamics

The response of the economy to a real shock, such as a change in productivity, the terms of trade or the burden of taxation, can be represented by a change in the variable q. It is clear from equations (6) and (6′) that the response of money wages and unemployment to a change in q, if it enters the two equations symmetrically, will be the same as the response to a demand shock, m. There has, however, been much discussion in the literature (e.g., Bruno and Sachs 1985; Grubb et al. 1983) of the idea that real shocks affect firms' pricing decisions but do not alter desired real wages in the wage bargain. Thus, for example, an adverse productivity or terms-of-trade shock might shift the equilibrium unemployment locus from u_0u_0 to u_1u_1 in Fig. 1, while leaving the equilibrium wage locus unchanged at w_0w_0. The economy would then move to a new equilibrium at point D, with the unemployment rate given by

$$u = \overline{u} - \frac{\Delta q}{\alpha_1 + \alpha_2} \tag{7}$$

where Δq is the change in productivity. It will be noted from the figure that a fall in productivity may in these circumstances raise money wages. The reason is that a fall in q raises costs and hence

prices, and increased prices will tend to raise money wages. Money wages will rise as long as the price effect outweighs the wage-depressing effect of higher unemployment.

The 1970s were characterized by particularly severe adverse supply shocks, in particular the oil price increases of 1973 and 1979 and slowdown of productivity growth throughout the industrialized world. The above analysis suggests that the capacity of an economy to adjust to such shocks will depend above all on the extent to which wage claims are moderated. Empirically there appears much support for the view that the more 'corporatist' the structure of wage bargaining in the economy (i.e., the more centralized the wage bargain) the more quickly are such supply shocks reflected in wage settlements (Bruno and Sachs 1985, ch. 11). Austria and Sweden are cited as examples of countries where the wage bargain is struck at the national level, involving centralized unions covering the bulk of the labour force, employers' associations and government. Corporatism is seen as helpful to the rapid assimilation of productivity changes and the like into the wage bargain both because it focuses attention on macroeconomic performance and because it avoids inter-union rivalry. In a decentralized system, individual wage bargainers may know about their individual sector but not about general macroeconomic developments, and may therefore be slow to adjust to macroeconomic shocks. Each group is reluctant to change its own wage if it is uncertain whether others will follow, because of concern over relative wages (Taylor 1980). Wage flexibility thus suffers from the 'paradox of isolation': each group might like to adjust its wage if it could be sure that similar adjustments would be made throughout the economy, but in a decentralized system there is no coordinating mechanism.

Finally, it may be noted that a rigidity of nominal wage rates, although it raises the unemployment costs of demand deflation, reduces the short-run costs of supply shocks. In Fig. 1, the progress of the economy from point *A* to point *D* is made slower if money wages are slower to adjust, and hence unemployment takes longer to emerge. A supply shock will raise prices and, if money wages are inflexible, the increase in prices will reduce real wages and thereby maintain employment. In this sense, real wage flexibility may be seen as the opposite of nominal wage flexibility (Sachs 1979). A number of authors (Bruno and Sachs 1985) have attributed the relatively strong performance of the United States economy since 1973 to a combination of a very high degree of nominal wage inflexibility (resulting in part from long-term wage contracts) and, over much of the period, demand expansionary policies. By contrast, in economies with more flexible money wages, meeting supply contraction by demand expansion would simply add faster inflation to higher unemployment.

See Also

- ▶ Trade Cycle
- ▶ Wages, Real and Money

Bibliography

Barro, R.J., and H.I. Grossman. 1971. A general disequilibrium model of income and employment. *American Economic Review* 61(1): 82–93.

Bruno, M., and J. Sachs. 1985. *Economics of worldwide stagflation*. Oxford: Basil Blackwell.

Fischer, S. 1977. Long-term contracts, rational expectations and the optimal money supply rule. *Journal of Political Economy* 85(1): 191–205.

Friedman, M. 1968. The role of monetary policy. *American Economic Review* 58: 1–17.

Grubb, D., R.A. Jackman, and R. Layard. 1983. Wage rigidity and unemployment in OECD countries. *European Economic Review* 21(1–2): 11–39.

Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan.

Leijonhufvud, A. 1968. *On Keynesian economics and the economics of Keynes*. New York: Oxford University Press.

Malinvaud, E. 1977. *The theory of unemployment reconsidered*. Oxford: Basil Blackwell.

Okun, A. 1975. Inflation: Its mechanics and welfare costs. *Brookings Papers on Economic Activity* 2: 351–401.

Sachs, J. 1979. Wages, profits and macroeconomic adjustment: A comparative study. *Brookings Papers on Economic Activity* 2: 269–193.

Taylor, J.B. 1980. Aggregate dynamics and staggered contracts. *Journal of Political Economy* 88(1): 1–23.

Wage Fund Doctrine

C. A. Blyth

A central part of classical analysis and closely related to the advances theory of capital, this doctrine lost support in the 1870s because of its association with unacceptable ideas on wages and trade unions. This loss was reinforced by J.S. Mill's authoritative 'recantation'. However, the doctrine was reaffirmed by Jevons and Böhm-Bawerk and survived at a high level of abstraction in neoclassical capital and production theory. This essay starts with the classical statement of J.S. Mill (1848), notices the recantation in 1869, and then looks both backwards to the eighteenth-century origins of the theory, and forwards to its post-classical developments.

Capital, says Mill, is a stock, previously accumulated, of the products of former labour. Because production takes time between the employment of labour and natural agents and the availability of their product, capital provides the shelter, protection, tools and materials which the work requires, and feeds and otherwise maintains the labourers during the process.

> Wages, then, depend mainly upon the demand and supply of labour; or as it is often expressed, on the proportion between population and capital. By population is here meant the number only of the labouring class, or rather of those who work for hire; and by capital only circulating capital, and not even the whole of that, but the part which is expended in the direct purchase of labour. To this, however, must be added all funds which, without forming a part of capital, are paid in exchange of labour, such as the wages of soldiers, domestic servants, and all other unproductive labourers. There is unfortunately no mode of expressing by one familiar term, the aggregate of what has been called the wages-fund of a country: and as the wages of productive labour form nearly the whole of that fund, it is usual to overlook the smaller and less important part, and to say that wages depend on population and capital. It will be convenient to employ this expression, remembering, however, to consider it as elliptical, and not as a literal statement of the entire truth.
>
> With these limitations of the terms, wages not only depend upon the relative amount of capital and population, but cannot, under the rule of competition, be affected by anything else. Wages (meaning, of course, the general rate) cannot rise, but by an increase of the aggregate funds employed in hiring labourers, or a diminution in the number of the competitors for hire; nor fall, except either by a diminution of the funds devoted paying labour, or by an increase in the number of labourers to be paid (Mill [1848], 1965, pp. 337–8).

This statement of the doctrine, agreeing in essentials with the views of Mill's contemporaries, for example, McCulloch and Senior, is followed by the conclusion that high wages require restraints on population growth.

In the recantation contained in his 1869 *Fortnightly Review* article on his friend Thornton's book, *On Labour*, Mill repeats the doctrine (Mill [1869], 1967, pp. 643–4) only to reject it immediately as a 'true representation of the matter of fact'. His grounds are simply that at any time the limit to the fund available to pay wages is not in practice fixed, because it includes 'the aggregate means of the employing classes'. The limit to the rise in wages is set by how much would drive the employer out of business. In the first six editions of his *Principles* Mill had said that if combinations of workmen 'aimed at obtaining actually higher wages than the rate fixed by supply and demand – the rate which distributes the whole circulating capital of the country among the entire working population – this could only be accomplished by keeping a part of their number permanently out of employment'. In the seventh edition (1871) the first part of this was replaced by '[workmen] would also have a limited power of obtaining, by combination, an increase of general wages at the expense of profits. But the limits of this power are narrow; and were they to attempt to strain it beyond those limits, this could only be accomplished [etc.]' (Mill [1871], 1965, p. 930). The explanation and political significance of Mill's disavowal are still the subject of academic debate; see, for example, Hollander (1985).

The idea of capital as a wages fund arises from the idea of capital as an advance to sustain labour during the period output takes to fructify, or be produced. The theory emerges naturally in

commercially agricultural and mercantile economies, and became a foundation stone of the developing economics of the late eighteenth century in Western Europe. The idea was expressed by Cantillon, Quesnay and Hume, and developed more fully by Turgot and Smith.

In prefacing his discussion of capital, or stock, Smith (1776) observes that the division of labour presupposes a previous accumulation of stock of provisions, materials and tools. He asserted that as the division of labour increases, in order to maintain constant employment, not only an unchanged stock of provisions but also a growing stock of materials and tools is necessary (Smith [1776], 1976, pp. 276–7). Smith's emphasis on the prior accumulation of this growing stock of materials and tools as a precondition of growth of employment, together with his attention to the dependence of wages on the relative rates of growth of capital stock and employment, is the foundation of all later discussion, especially of the British school.

Furthermore, Smith recognized that although the growth of capital could outstrip the growth of population, thus allowing wages to rise for lengthy periods, population growth being responsive to wages above a subsistence level could in the long run reduce wages to that level. This even-handed recognition of various possibilities also influenced later discussion. However, Malthus's contribution, popularly interpreted as a prediction of an inevitable approach of wages to near starvation levels, coupled with the acceptance of a fixed wage fund, provided the basis in the early ninteenth century of pro-employer journalism and other advocacy which was both dismal and anti-trade union. The 'iron law' of Lassalle and later socialists stemmed more from popular Malthusianism than from the wage fund, and the majority of British economists, however dismal they may have appeared, cannot be accused of Malthusian predictions and pro-employer advocacy.

Wage fund doctrine, in which the wage rate is determined by supply and demand for labour, the demand for labour depending on the size of the wage fund, may be and has been interpreted variously, both in the early ninteenth century and today. How wide the net of the doctrine is cast, and whether it is co-extensive with the advances theory of capital, are largely matters of intent. Ricardo, for instance, defines the natural price of labour as the price necessary to enable workers 'to subsist and to perpetuate their race, without either increase or diminution' (Ricardo [1817], 1951, p. 93). Following Smith, he envisages differences between the market price and the natural price to cause changes in the labour force (through population change) which eventually equate market and natural prices at zero labour force change. Thus, if the attention is focused on natural price Ricardo (like Marx) may be said to hold a theory in which the wage rate is exogenous at a given subsistence level. However, if attention is focused on market price, Ricardo adheres to a distinctly Smithian wage fund doctrine, with increases in capital giving rise to increased demand for labour, and consequential changes in market price and in labour supply performing equilibrating functions which may be so slow that, as Ricardo says, they may be overtaken by fresh increases in capital. Failure to recognize both views in Ricardo has led some modern commentators to contrast Ricardo, holding the former view, with the anti-Ricardians, holding the latter view.

Regarding the statement that the ratio of wage fund to labour force determines the wage rate as largely self-evident, the discussion of most early ninteenth-century economists centred on the equilibrating effect of changes in market rate on labour supply, and the relation of wage fund to stock of capital. Malthus (1820) and McCulloch (1825), for instance, state quite clearly that because the labour supply cannot quickly adjust to changes in the market rates (McCulloch refers to a delay of 18–20 years), the natural rate will move to some extent with the market rate, good times allowing workers to raise their standards, bad times forcing them to lower them. There was general agreement with Ricardo that the equilibrating mechanism would work only slowly or weakly, although there were differences, Malthus, for instance, preferring to replace Ricardo's definition of natural price (because it presupposes a stationary state) by a definition which requires 'an average supply of labourers'.

The discussion of fixity of the wage fund brought forth a variety of ideas. Malthus argued that a change in the wage fund does not necessarily imply a change in the demand for labour, giving several reasons: for example, a fall in the price of raw produce, if sudden, may because of general distress reduce the demand for labour; while an increase in circulating capital for the production of luxuries may not increase the demand for labour (Ricardo agreed with him). However, Malthus believed that increased use of fixed capital was associated with increased use of circulating capital and increased demand for labour, provided the market for the produce increased in proportion. This relatively subtle analysis of the relation of capital and labour in production was quite common. Senior (1836), for instance, who like McCulloch supported assisted emigration and deprecated wage subsidies on liberal Malthusian grounds, distinguishes the wage fund from capital, interpreting the former as a stock of means of production to supply workers with their future needs. The size of the fund depends on the productiveness of labour (which in turn depends on the use of capital), and on the extent to which labour is diverted from the production of wage goods, for example, by high levels of rent, taxation and profit. The determinants of profit lead Senior to a discussion of the length of the period of advances which foreshadows Jevons and Böhm-Bawerk.

Obviously, despite inconsistencies and the problems of interpretation, classical wage fund doctrine could be rich and sophisticated, although historians of the doctrine (Taussig 1896; Blaug 1958) have argued that views about the fund were both varied and vague. Nevertheless, mid-century critics like Longe and Thornton certainly fastened their attack on the fixity of the wage fund, and J.S. Mill was sufficiently embarrassed by the attack to accept the criticism. Part of the problem was the undeveloped analysis of substitution between fixed capital, raw materials and labour (the first two being usually simply subtracted from total capital leaving a wage fund as a residual). Cairnes, who was unhappy with the recantation, attempted a reconciliation (1874) by contrasting the ratio (presumably given for each industry) of capital to labour with the ratio of capital to wage fund. A change in the supply of labour, causing the wage rate to alter, would alter the distribution of labour between industries and change the size of the wage fund.

In the subsequent evolution of capital and production theory, the advances theory was further developed by Jevons and Böhm-Bawerk. Jevons (1871), while rejecting the wage fund doctrine as a truism, defined capital as a subsistence fund with a time dimension, and the Austrian, Böhm-Bawerk (1889), contributed the seminal idea of a variable period of production (during which advances are made) whose optimal length is chosen by the capitalist. In this, Böhm-Bawerk may have been anticipated by Ricardo. Wicksell (1893) interpreted Böhm-Bawerk and Jevons in a general equilibrium framework, explaining capital as in essence a fund of the produced means of production, but including exhaustible resources and excluding permanent improvements (i.e. broadly containing Smith's provisions, materials and tools), whose value in a stationary state is approximated by the annual wage bill and payments for primary resources, plus the interest accumulated over the period of production.

Thus at a high level of abstraction the classical dictum that the wage rate is determined by the ratio of wage fund to labour force was replaced by a typical neoclassical equilibrium relationship: the value of capital per unit of labour equals (approximately for low rates of interest) the wage rate multiplied by the period of production. Later writers on capital (e.g. Hayek 1941) avoided the concept of period of production, for the same reason that the most recent theorists have found it difficult to use the concept of capital itself (Pasinetti 1977).

However, despite these modern theoretical developments which have diluted simple ideas, it is possible that in much of the modern underdeveloped world the simple eighteenth-century idea of the wage fund, like the equally simple ideas of the quantity theory of money and the labour theory of value, may retain practical relevance.

See Also

- ▶ British Classical Economics
- ▶ Longe, Francis David (1831–c1905)
- ▶ Mill, John Stuart, as Economic Theorist

Bibliography

Blaug, M. 1958. *Ricardian economics*. New Haven: Yale University Press.

Böhm-Bawerk, E. von. 1889. *Positive theory of capital*. South Holland: Libertarian Press, 1959.

Cairnes, J.E. 1874. *Some leading principles of political economy and their applications*. London: Macmillan.

Hayek, F.A. 1941. *The pure theory of capital*. London: Routledge.

Hollander, S. 1985. *The economics of John Stuart Mill*. Oxford: Blackwell.

Jevons, W.S. 1871. *The theory of political economy*. 3rd edn, London: Macmillan, 1888.

Malthus, T.R. 1820. *Principles of political economy*. New York: Augustus M. Kelley, 1968.

McCulloch, J.R. 1825. *The principles of political economy*. London: Murray, 1870.

Mill, J.S. 1848. *Principles of political economy*. In *The collected works of John Stuart Mill*, Vols II and III (variorum edition), Toronto: University of Toronto Press, 1965.

Mill, J.S. 1869. Thornton on labour and its claims. In *The collected works of John Stuart Mill*, Vol. V, Toronto: University of Toronto Press, 1967

Pasinetti, L.L. 1977. *Lectures on the theory of production*. New York: Columbia University Press.

Ricardo, D. 1817. On the principles of political economy and taxation. In *The works and correspondence of David Ricardo*, Vol. I, ed. P. Sraffa, Cambridge: Cambridge University Press, 1951.

Senior, N. 1836. *An outline of the science of political economy*. London: Allen & Unwin, 1938.

Smith, A. 1776. An inquiry into the nature and causes of the wealth of nations. Vol. I, Oxford: Clarendon Press, 1976.

Taussig, F.W. 1896. *Wages and capital*. London: London School of Economics and Political Science, 1935.

Wicksell, K. 1893. *Value, capital and rent*. London: Allen & Unwin, 1954.

Wage Goods

Ingrid H. Rima

The concept of 'wage goods' had socio-political as well as economic significance in the early history of distribution theory. The 'necessaries' that comprised the subsistence of workers set them apart, as a class, from landowners and capitalists who consumed luxuries and conveniences.

Important pre-Smithian arguments were predicated on the wage goods concept. Mercantilist writers, while chiefly concerned with augmenting national gold stocks, attributed the superiority of advanced countries to the new technologies which they were able to support with their large stocks of wage goods. These goods originated as surpluses produced by farm workers and were advanced to 'free hands' by merchant capitalists. Most writers urged that the quantity of wage goods paid be limited to subsistence needs on the premise that 'idleness will not be totally rooted out, until people are forced in one way or another to give up both superfluity and days of recreation' (Steuart, *Inquiry II*, p. 691).

Beginning with Adam Smith, classical writers incorporated the wage goods concept into more sophisticated theoretical analyses. For example, Book I of *The Wealth of Nations*, makes it clear that 'capitalistic' production, which is roundabout and time consuming but more productive than direct methods, is possible only because capital in the form of materials and wage goods is available to be advanced to labour. As Smith observed (Introduction: Book II):

> A weaver cannot apply himself entirely to his peculiar business, unless there is beforehand stored up somewhere, either in his own possession or in that of some other person, a stock sufficient to maintain him and to supply him with the materials and tools of his work, till he has not only completed, but sold his web. This accumulation must, evidently, be previous to his applying his industry for so long a time to such a peculiar business.

Smith interprets wage goods more broadly than earlier writers did by including in the concept not only worker 'necessaries' but also 'conveniences of life'. His rudimentary 'wage fund' theory interpreted the stock of wage goods as constituting a demand for labour. Coupled with the simplistic assumption that labour supply is equal to the country's total population, it led to the thesis that the general real wage level in the economy reflects the ratio between the stock of wage goods and the population. As the real

counterpart of money wages, the wage goods concept also became the basis for Smith's theory of the relationship between corn prices and money wages. Because corn was the chief article of labour's subsistence, Smith thought the worker's money wage to be more dependent on the average price of corn than it is on the price of meat or rude produce generally (1776, p. 187). The chapter entitled 'Of Wages' explains both the money (or nominal) price of labour and its price in corn, and argues that corn prices are the regulator of all other commodity prices (1776, pp. 476–7).

The notion that *parsimony* or *abstinence* from consumption is a prerequisite for the production of wage goods (or their monetary equivalent, *the wage fund*) is particularly relevant because it sanctions capitalist profits as a separate class income. Profit came to be viewed as an *earned* and, therefore, proper income share quite early in the history of economics, precisely because the dependence of workers on masters who advance them both maintenance and the materials of their work was recognized.

David Ricardo made the concept of 'wage goods' central to the problem of explaining the distributive (or income) shares, on the premise that these enter into the production of *all* products while non-wage (luxury) goods do not. Ricardo's 'fundamental theorem of distribution' which turns on a rigid distinction between wage goods and luxury goods, implies that the general rate of profits and thus of all income shares, depends chiefly on conditions of production in the wage goods (i.e. 'corn') sector and the real wage.

On occasion, both Ricardo and Thomas Malthus distinguished between 'absolute necessaries' and wage goods (Ricardo letter dated 29 November 1820, *Works* Vol. VIII, p. 311) in examining the components of capital and the ability of income receivers to pay taxes and/or support savings. It is in this context that he noted that more is generally allotted to the labourer under the name of wages than is sufficient for maintaining him and his family (*Works*, I, pp. 421–2).

Modern writers have generally dismissed the classical distinction between wage goods and non-wage goods on the ground that all commodities are consumed by workers and non-workers. Nevertheless, the wage goods concept was revived by A.C. Pigou in *Industrial Fluctuations* (1927, pp. 115–16) and *Theory of Unemployment* (1933, Part 1, chs 4 and 5). His 1933 work introduces the notion of the representative wage earner who buys both necessaries and luxuries. The proportions of purchase may vary if tastes change or relative prices become altered by changes in productive technique or the demands of non-wage earners for goods (pp. 12–18). Analogously, alterations in money wages, other things remaining constant, may alter the proportions in which the representative wage earner makes his purchases. These facts make it difficult to determine precisely what a wage good unit is in given circumstances (pp. 18–20).

J.M. Keynes also made important use of the wage goods concept in his *General Theory of Employment Interest and Money* (1936). His concern was that inflationary bank policy, in raising the average price of wage goods and thereby reducing the *real* wage rate, would be unable to assure that all labour markets would clear. Keynes's argument was that bank policy can satisfy 'the first postulate of classical theory' by increasing the demand for labour, but it cannot fulfill the second postulate which requires that the labour supply decrease sufficiently to clear the market because the real wage rate (i.e. the price of wage goods) may not measure the disutility of labour (Keynes 1936, chs 2 and 19, Appendix A; Rima 1984). Failure of the market to satisfy the second postulate underlies Keynes's notion of involuntary unemployment.

Bibliography

Keynes, J.M. 1936. The general theory of employment, interest and money. (Reprinted in *The collected writings of John Maynard Keynes*, vol. VII. London: Macmillan, 1973).

Pigou, A.C. 1927. *Industrial fluctuations*. London: Macmillan.

Pigou, A.C. 1933. *The theory of unemployment*. London: Macmillan.

Rima, I. 1984. Involuntary unemployment and the respecified labor supply curve. *Journal of Post-Keynesian Economics* 6(4), Summer: 540–550.

Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*. New York: Modern Library, 1937.
Sraffa, P. (ed.). 1951–2. *Works and correspondence of David Ricardo*, vols I and VIII. Cambridge: Cambridge University Press.
Steuart, Sir J. 1767. *An inquiry into the principles of political economy*. In ed. A.S. Skinner. Chicago: University of Chicago Press, 1966.

Wage Indexation

Joshua Aizenman

Keywords
Consumer price index; Contingency contracts; Inflation; Monetary policy rules; Nominal wage contracts; Rational expectation; Staggered wage setting; Wage flexibility; Wage indexation

JEL Classifications
J3

Wage indexation is a mechanism designed to adjust wages to information that cannot be foreseen when the wage contract is negotiated. A wage contract with indexation clauses will specify the wage base (that is, the money wage applicable in the absence of new information), the indexation formula that will be used to update wages, and how often updating will occur. Most traditional discussion has focused on wage indexation to the price level as a mechanism to stabilize real wages in the presence of inflation. More recently, however, attention has shifted to indexation to a wider set of indicators. These indicators include both richer price information (such as the value added price deflator and the terms of trade) and rules designed to index wages to indicators measuring the level of nominal activity (such as nominal GNP). Concurrently, growing attention has been given to the potential role of wage indexation in affecting the will and ability to reduce inflation.

The economic evaluation of the role and desirability of wage indexation is inherently tied to the assessment of the functioning of the labour market and the role of wage contracts. If labour markets are cleared continuously, as in an auction market (that is, if wages are adjusted continuously to equate the demand and the supply of labour), wage contracts serve no purpose and indexation clauses are either redundant or diminish welfare. On the other hand, the potential role of various wage indexation schemes grows the further we move away from an auction labour market. Consequently, an analysis regarding the role of wage indexation invites a specification of the nature of the deviations from an auction labour market and of the disequilibrium mechanism applied in that market. Indeed, challengers of the usefulness and relevance of wage indexation have remarked on the lack of rigorous understanding of the postulated deviations from auction market behaviour (see Barro 1977). At the same time, a growing body of research has proceeded on the assumption that the existence of nominal contracts with limited degrees of indexation provides enough evidence to reject an auction labour-market clearing hypothesis (see Fischer 1977b). This assumption has justified studies of the economics of wage indexation in models that lack a rigorous general equilibrium framework, but still provide insights into complicated economic environments. We start with a review of analytical studies on wage indexation, continue with overview of experience with wage indexation in various countries, and close with some interpretative remarks.

Analytical Aspects of Wage Indexation

The usefulness of indexing wages to the price level has been the subject of considerable research, and perceptive comments on the topic can be found in publications going back to Keynes's *General Theory of Employment, Interest and Money* (1936). A renewed interest in the question was generated by Gray (1976, 1978) and Fischer (1977a), who integrated the rational expectation hypothesis with nominal contracts. They considered an economy where nominal

contracts preset the contract wage before the realization of stochastic shocks. The rational expectation hypothesis is invoked to determine the contract wage, which is set at a level that is expected to clear the labour market. The contract agreement also specifies the degree of wage indexation to unanticipated inflation. A complete indexation implies real wage rigidity, whereas the absence of indexation entails nominal wage rigidity in which changes in the price level directly affect the real wage. The contract specifies also the determination of employment rule, which is assumed to be demand determined. Consequently, in general employment will deviate from the flexible equilibrium level (that is, from the employment level that will prevail in an economy where the wage is set as to clear the labour market continuously). The optimal degree of indexation is designed so as to minimize the expected squared output deviations from its market clearing level. This can be shown to be equivalent to minimizing the deadweight loss in the labour market for risk-neutral agents (see Aizenman and Frenkel 1985). The optimal degree of wage indexation is a compromise between two opposing forces: the wish to neutralize the potential output consequences of monetary (nominal) shocks by keeping real wages stable, and the wish to reduce real wages in the presence of adverse real shocks. The first goal is accomplished by complete wage indexation to prices, and the second by partial indexation. Optimal indexation balances between these two forces, implying that greater importance of monetary relative to real shocks will be associated with higher indexation. Such an indexation scheme implies that the real sector is not insulated from monetary variability (see Gray 1976; Fischer 1977a). As a result, optimal indexation will tend to stabilize output around its full equilibrium level while it will tend to increase the volatility of prices.

Subsequent research had raised several important questions, for instance why wages are contingent only on prices and not on other relevant information. As Barro (1977) and Karni (1983) have pointed out, optimal contingencies will allow wages to adjust to all relevant information, thereby clearing the labour market continuously and eliminating the output effects of monetary policy. The fact that we find no contracts with rich sets of contingencies suggests, however, that it will be very costly to collect and process all the information needed to write and enforce full contingency contracts (see Fischer 1977b; Blanchard 1979). Another related question is the underlying justification of the disequilibrium hypothesis. As demonstrated by Cukierman (1980), the indexation derived by Gray is affected by the disequilibrium hypothesis. It can be shown, however, that this issue becomes inconsequential once we approach a full contingency contract because such a contract will clear the market independently of the disequilibrium hypothesis (see Aizenman and Frenkel 1985).

Further developments regarding wage indexation have extended the analysis to open economies. Flood and Marion (1982) showed that optimal indexation is determined by the exchange rate regime, whereas Aizenman and Frenkel (1985) demonstrated that optimal indexation is only one among many potential policies, and that there is a close linkage between wage indexation, monetary policy and exchange rate policies.

A relevant consideration in these discussions is the set of indicators to which the wage is indexed. Most of the above studies derived optimal indexation rules in terms of the underlying structural parameters (like the elasticities of demand in the money and labour market). While these results are informative, their usefulness is limited by the degree of availability of information regarding the underlying structure. In an environment with limited and costly information, indexation rules that use easily available data, without relying on the structural parameters, should have natural advantage. Such rules were studied by Marston and Turnovsky (1985) who investigated the usefulness of wage indexation to the GNP price deflator and to the GNP in the context of energy shocks. Aizenman and Frenkel (1986) pursued related research, showing that if the elasticity of demand for labour exceeds the elasticity of supply, then indexing nominal wages to nominal GNP is preferable to indexing to the value added price index, and this, in turn, is preferable to indexation to the CPI (this ranking is reversed

when the elasticity of the supply of labour exceeds the elasticity of demand). Similar results are applicable for ranking the usefulness of targeting monetary policy to the above indicators. Taking another research direction, Fethke and Policano (1984) addressed the usefulness of coordinating the timing of wage negotiations in a multisectoral economy. They concluded that when disturbances are driven primarily by relative shocks (that is, shocks that hit the two sectors differentially) staggered negotiation is optimal.

Once we place wage indexation in its proper perspective as a macroeconomic policy instrument, a natural question arises regarding the linkages between wage indexation and other policies such as taxes and assets indexation (see Friedman 1974), the risk-sharing effects of indexation (see Azariadis 1978) and wage renegotiation (see Gray 1978; Aizenman 1984). Further analysis and references regarding these important topics can be found in a useful conference volume (Dornbusch and Simonsen 1983).

Experience with Wage Indexation

The experience with indexation of the last decades has been mostly with various degrees of wage indexation to the CPI. The precise indexation policy differs across countries considerably, depending on the centralization of the wage negotiation process and the degree to which wage indexation is viewed as income policy instead of as an instrument to enhance the efficiency of the labour market. For example, in the United States wage indexation is allowed, but there are no guidelines and the details of the indexation schemes are left for the contract negotiation. In Europe, Latin America and Israel labour negotiation tends to be more centralized, and the indexation provisions tend to be dictated by a centralized policymaker. Some countries (for instance, Italy and Brazil) have applied wage indexation as an implicit income policy. This was done by imposing a rigid base wage and a high degree of wage indexation (and in some cases with a cap at high income levels). Such a policy is a poor substitute for direct income policy because it generates distortions in the labour market. These distortive effects increase in periods associated with real shocks, such as changes in input prices and in aggregate demand. Other countries have attempted to design partial indexation as a device to allow real wage flexibility in the presence of terms of trade shocks (see Brenner and Patinkin 1977).

While experience with indexation differs across countries, several observations appear to be common to them all. *First*, the degree of indexation to the price level and the frequency of wage adjustment tend to go up with the level and volatility of inflation (see Ehrenberg et al. 1983; Kleiman 1977). *Second*, a higher indexation rate tends to reduce linkages between excess demand forces in the labour market and wages (see Sachs 1983). *Third*, limited indexation seems not to be a controversial issue for countries with stable and relatively low inflation rates. For countries with high and volatile inflation the desirability of wage indexation is an important policy issue when attention shifts to curbing that inflation. In various countries in the last decades we have observed the adoption of indexation at low and moderate inflation rates. Once inflation has risen to intolerable levels, however, the policymaker has tended to couple abrupt disinflationary policies with disindexation policies (for example in Iceland in 1983 and in Israel in 1985). This tendency is related to the fact that a typical indexation scheme adjusts wages to lagged inflation, implying that it builds in inertia, thereby a policy of disinflation will tend to raise real wages during the transition, generating unemployment (see Simonsen 1983; Fischer 1984).

Concluding Remarks

The role of nominal contracts and the potential role of wage indexation and macro policies are major research topics in macroeconomics. In recent years we have witnessed considerable development in this area, achieved by integrating the rational expectation hypothesis into models where transaction costs prevent continuous auction market clearing. The present state of

theoretical research is, however, far from satisfactory. On the one hand, the theoretical papers reviewed above do not offer a framework that will satisfy 'purists', although they allow assessment of important policy issues in the presence of realistic contracts. On the other hand, 'purists' have not so far been able to explain the existence of nominal contracts of the type observed in various segments of the labour market. Interesting research directions that may provide further clues are frameworks that will recognize and model economic environments where decisions are costly. These costs stem from the observation that data gathering and screening are not free, and that resources are lost in the negotiation process. Such a framework will put a premium on simpler decision rules requiring less frequent negotiation, and nominal wage contracts may be one important example of such rules. The research into nominal contracts reviewed above is, we may hope, a step in that challenging direction.

The experience with wage indexation to prices suggests that greater attention should be given to the design of tractable indexation rules that will generate real wage flexibility in the presence of real shocks while retaining the purchasing power of wages in the presence of nominal disturbances. Such rules should be based upon widely available information. A candidate that deserves further exploration is wage indexation to nominal GNP. Simple-minded rules for indexation to the CPI have several potential disadvantages. They prevent real wage and employment adjustment in the presence of real shocks, thereby causing suboptimal employment. In the presence of nominal shocks and inflation, indexation to prices can generate dynamic inconsistencies – in the short and intermediate run it mitigates the losses associated with unanticipated inflation, but it thereby reduces the will to follow policies that are prudent with regard to inflation, causing higher inflation in the long run. Once the policymaker attempts to disinflate, the indexation scheme may exacerbate the welfare costs associated with the transition to lower inflation. Thus, a policy device that is viewed as useful in the short run can be harmful in the long run.

Consequently, indexation rules are not a substitute for prudent macro-policies. Rules that index wages to nominal income or to the GDP deflator can serve a useful role as part of macro-policies that recognize the need to undergo real adjustment in the presence of real shocks. At the same time, they are deceptive and harmful if they are used as income policy tools in an attempt to maintain the purchasing power of wages in economies exposed to productivity and terms-of-trade shocks.

Bibliography

Aizenman, J. 1984. Optimal wage re-negotiation. *Journal of Monetary Economics* 13: 251–262.

Aizenman, J., and J.A. Frenkel. 1985. Optimal wage indexation, foreign exchange intervention, and monetary policy. *American Economic Review* 75: 402–423.

Aizenman, J., and J.A. Frenkel. 1986. Supply shocks, wage indexation, and monetary accommodation. *Journal of Money, Credit and Banking* 18: 304–322.

Azariadis, C. 1978. Escalation clauses and the allocation of cyclical risks. *Journal of Economic Theory* 18 (1): 119–155.

Barro, R.J. 1977. Long-term contracting, sticky prices, and monetary policy. *Journal of Monetary Economics* 3: 305–316.

Blanchard, O.J. 1979. Wage indexation rules and the behavior of the economy. *Journal of Political Economy* 87: 798–815.

Brenner, R., and D. Patinkin. 1977. Indexation in Israel. In *Inflation theory and anti-inflation policy*, ed. E. Lunberg. London: Macmillan.

Cukierman, A. 1980. The effects of wage indexation on macroeconomic fluctuations. *Journal of Monetary Economics* 6: 147–170.

Dornbusch, R., and M.H. Simonsen, eds. 1983. *Inflation, debt and indexation*. Cambridge, MA: MIT Press.

Ehrenberg, R., L. Danziger, and G. San. 1983. Cost of living adjustment clauses in union contracts: A summary of results. *Journal of Labor Economics* 1: 215–245.

Fethke, G., and A. Policano. 1984. Wage contingencies, the pattern of negotiation and aggregate implications of alternative contract structure. *Journal of Monetary Economics* 14 (2): 151–170.

Fischer, S. 1977a. Wage indexation and macro-economic stability. In *Stabilization of domestic and international economy*, ed. K. Brunner and A.H. Meltzer. *Journal of Monetary Economics*, Supplementary 5, 107–147.

Fischer, S. 1977b. Long term contracting, sticky prices, and monetary policy: A comment. *Journal of Monetary Economics* 3: 317–323.

Fischer, S. 1984. *Real balances, the exchange rate and indexation: Real variables in disinflation*, Working Paper No. 1497. Cambridge, MA: NBER.
Flood, R.P., and N.P. Marion. 1982. The transmission of disturbances under alternative exchange-rate regimes with optimal indexing. *Quarterly Journal of Economics* 97: 43–66.
Friedman, M. 1974. *Monetary correction*. Washington, DC: American Enterprise Institute.
Gray, J.A. 1976. Wage indexation: A macro-economic approach. *Journal of Monetary Economics* 2: 221–235.
Gray, J.A. 1978. On indexation and contract length. *Journal of Political Economy* 86: 1–18.
Karni, E. 1983. On optimal wage indexation. *Journal of Political Economy* 91: 282–292.
Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan.
Kleiman, E. 1977. Monetary correction and indexation: The Brazilian and Israeli experience. *Explorations in Economic Research* 4 (1): 141–176.
Marston, R.C., and J.S. Turnovsky. 1985. Imported material prices, wage policy, and macroeconomic stabilization. *Canadian Journal of Economics* 18: 273–284.
Sachs, J. 1983. Real wages and unemployment in the OECD countries. *Brookings Papers on Economic Activity* 1983 (1): 255–304.
Simonsen, M.H. 1983. Current theory and the Brazilian experience. In *Inflation, debt, and indexation*, ed. R. Dornbusch and M.H. Simonsen. Cambridge, MA: MIT Press.

Wage Inequality, Changes In

Stephen Machin and John Van Reenen

Abstract

We examine trends in wage inequality in the United States and other countries since the 1960s. We show that there has been a secular increase in the 90–50 wage differential in the United States and the United Kingdom since the late 1970s. By contrast the 50–10 wage differential rose mainly in the 1980s and flattened or fell in the 1990s and 2000s. We conclude that a version of the skill-biased technical change hypothesis combined with institutional changes (the decline in the minimum wage and trade unions) continues to offer the best explanation for the observed patterns of change.

Keywords

Decentralization; Elasticity of substitution; Heckscher–Ohlin trade theory; Imperfect competition; International trade; Labour market institutions; Minimum wages; Skill-biased technical change; Stolper–Samuelson theorem; Technical change; Trade unions; Unemployment; Wage differentials; Wage inequality

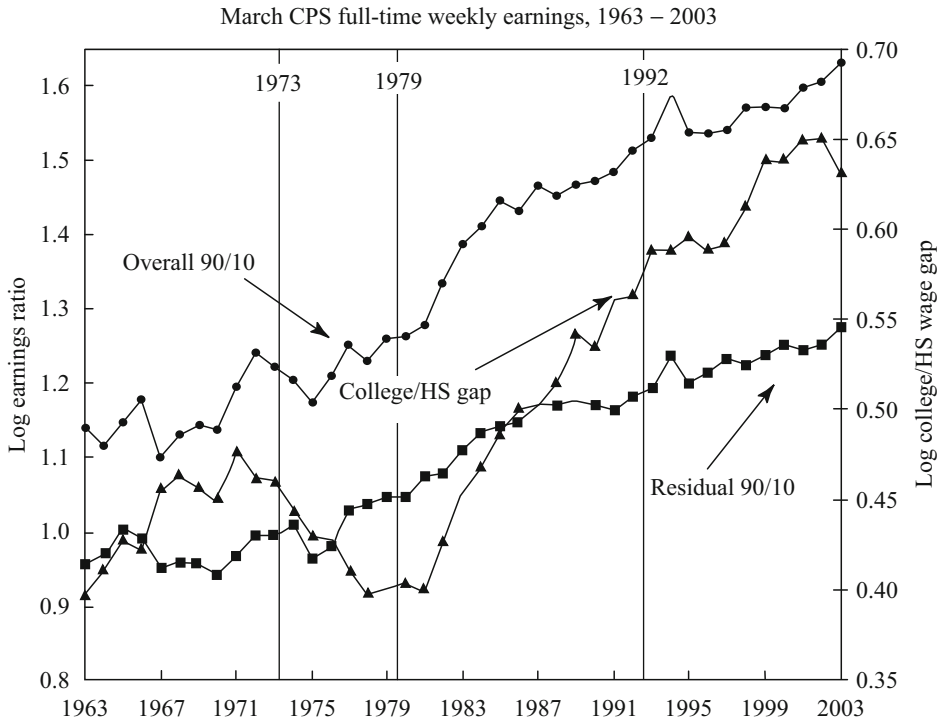
JEL Classifications

J300

Introduction

Study of the structure of wages has been a preoccupation of economists for a long time and dates back at least as far as Adam Smith. Until the early 1990s economists commented on the remarkable stability of the wage structure in the post-war period. But then many empirical studies (for example, Bound and Johnson 1992) noticed that wage inequality in America had risen dramatically since the late 1970s. Related empirical research (notably by Goldin and Katz 1999, 2001) went back further in time uncovering other periods of changing wage structures in American history. Other countries, notably the United Kingdom, also saw a significant increase in wage inequality at about the same time as the recent US changes (Machin 1996). These observations kick-started what has become a huge empirical and theoretical literature seeking to measure and explain changes in wage inequality (see the survey of Katz and Autor 1999). Since wages are a major part of people's income and economic well-being, the increase in wage inequality feeds through to income, consumption and poverty rates. So understanding the patterns of wage inequality is important from a normative as well as a positive perspective.

In this article we examine what has happened to the wage distribution since the 1960s, looking principally at the United States, where the bulk of the economic research has focused, but where possible also examining other countries. Section "What Has Happened to the Wage



Wage Inequality, Changes In, Fig. 1 Changes in US wage inequality, 1963–2003. *Note*: based on full-time weekly earnings for all workers in the March *Current Population Survey* (*Source*: Autor et al. (2005))

Distribution?" describes the observed changes in the structure of wages (although we fully acknowledge there are some contentious, and as of yet unresolved, issues about the observed patterns of change). Section "Explanations of Changes in Wage Inequality" looks at the main explanations of the observed changes that have emerged from the large body of work in this area. Section "Conclusions" offers some conclusions.

What Has Happened to the Wage Distribution?

Overall Trends in US Wage Inequality

To set the scene, Fig. 1 plots out the salient features of the US full-time weekly wage distribution from 1963 through to 2003. At least three things stand out from Fig. 1. First, educational wage differentials – measured as the gap in pay between college and high school educated workers – have risen consistently since 1979 (after falling somewhat in the 1970s and rising somewhat in the 1960s). The rate of increase was more rapid in the 1980s than after 1992. (This ongoing secular rise in educational wage premia is also seen in the hourly wage series from March outgoing rotation group of the Current Population Survey, CPS; see Lemieux 2006.) Second, the 90–10 wage differential – defined as the difference in weekly pay for those at the 90th and 10th percentiles of the overall wage distribution – has been rising since 1976 (and maybe even earlier). Third, the 'residual' 90–10 wage differential – the difference between those at the 90th and 10th percentiles of the overall wage distribution after controlling for education, experience and gender – has risen consistently since 1967, especially after the mid-1970s (see Juhn et al. 1993). This increase in 'within group' wage inequality has also generated much excitement and interest from theorists, but is particularly hard to interpret in the light of compositional changes (Lemieux 2006).

Even though different data-sets show some differences and there are some variations in inequality measures across data sources, the overall picture is one of a dramatic increase in American wage inequality since 1979.

Comparing the United Kingdom with the United States

The United Kingdom is another country where wage inequality has risen dramatically. Comparison of the United States and United Kingdom is useful to pin down certain issues to do with the rise in wage inequality. One important point is that since 1980 there are marked decadal differences in the opening up of the wage structure. Analysis of US and UK micro-data uncovers a clear picture for the 1980s in both countries: wage growth was more pronounced at higher points of the distribution, and faster wage growth higher up the distribution is almost monotonic in both counties, leading to large increases in wage inequality. An important difference, however, is that in the United Kingdom there was positive wage growth throughout the distribution whereas in the United States workers in the bottom quartile actually experienced zero or negative wage growth.

The picture becomes more complex post-1990. In both countries the 90–50 continues to diverge ('upper tail inequality') whereas the 50–10 ('lower tail inequality') in the United States actually shrinks, indicating some wage compression. In the United Kingdom the 50–10 is stable (increasing a bit in the 1990s and shrinking a bit in the 2000s). Overall then, the increase in wage inequality has been stronger in the upper tail than the lower tail taking the period as a whole, and has been more pronounced in the 1980s than post-1990.

A marked and important similarity between the two countries is the continuous and rapid growth of wages at the very top of the distribution. In the United Kingdom, wage growth at the 95th percentile (and above) is greater than at other percentiles of the wage distribution in the 1980s, 1990s and 2000s. This is also true for the United States (except for the 10th percentile in the 1990s). So within the picture of overall rising inequality the very rich have done particularly well.

Wage Inequality, Changes In, Table 1 Male 90–10 wage ratios across countries, 1980–2000

	Male 90–10 wage ratios		
	1980	1990	2000
Australia	2.73	2.71	3.16
Finland	2.44	2.57	2.47[f]
France	3.38	3.46	3.28[e]
Germany	2.53[b]	2.44	2.86[e]
Italy	2.09[b]	2.38	2.44[c]
Japan	2.60	2.84	2.74[f]
Netherlands	2.32[a]	2.48	2.83[f]
New Zealand	2.72	3.08	3.55[d]
Sweden	2.11	2.07	2.35[e]
UK	2.63[b]	3.24	3.40
US	3.58	4.41	4.76

Note: Data is from different years where indicated by the following superscripts: a – 1985; b – 1986; c – 1996; d – 1997; e – 1998; f – 1999

Source: OECD data website http://www.oecd.org

The other key feature of the changing wage distributions in the United Kingdom and the United States (and elsewhere) has been the polarization of work into 'good jobs' and 'bad jobs' (defined as high-wage and low-wage jobs). While there has been significant growth in well-paid 'good jobs' at the upper tail of the distribution (like lawyers, senior managers and consultants) there has been an increase in low-paid 'bad jobs' in the lower tail of the distribution (like cleaners, hairdressers, shop assistants and burger flippers). In the 1990s especially it seems that the middle of the distribution seemed to do somewhat worse than those at the top or bottom. These findings have been reported on in the United States (Autor et al. 2006), United Kingdom (Goos and Manning 2007) and Germany (Spitz-Oener 2006).

The Experience of Other Countries

There is less systematic evidence for the evolution of the wage distribution outside of the United States and the United Kingdom, especially for more recent years. Table 1 uses OECD data to show 90–10 male wage ratio for a range of countries between 1980 and 2000. Broadly speaking, the 1980s rise in inequality was seen only in the United Kingdom and the United States and in specific countries where particular episodes to

Wage Inequality, Changes In, Table 2 Aggregate trends in graduate/non-graduate employment and wages, UK and USA, 1980–2004

	UK		USA	
	% graduate share of employment	Relative weekly wage (full-time)	% graduate share of employment	Relative weekly wage (full-time)
1980	5.0	1.48	20.8	1.41
1985	9.8	1.50	24.2	1.53
1990	10.2	1.60	25.7	1.60
1995	14.0	1.60	31.8	1.65
2000	17.2	1.64	31.8	1.69
2004	21.0	1.64	34.2	1.66
Changes:				
1980–2004	16.0	.16	13.4	.25
1980–90	5.2	.12	4.9	.19
1990–2000	7.0	.04	6.1	.09
2000–4	3.8	.00	2.4	– .02

Notes: Sample is all people aged 18–64 in work and earning, except for relative wages, which are defined for full-time workers. The relative wage ratios are derived from coefficient estimates on a graduate dummy variable in semi-log earnings equations controlling for age, age squared and gender (they are the exponent of the coefficient on the graduate dummy)

Sources: UK: derived from General Household Survey (GHS) and Labour Force Survey (LFS); updated from Machin and Vignoles (2005). US: derived from Current Population Survey data

move to a much more market-oriented economy occurred (notably New Zealand). Elsewhere wage inequality did not alter much. The 1990s is a little different, with evidence of widening wage structures starting to occur in places previously characterized by stable wage structures – Germany is a very good example of this. Moreover, as we discuss below, the Continental European countries did experience a larger increase in unemployment, which may be due to the same underlying forces that have pushed up wage inequality in Britain and America.

Explanations of Changes in Wage Inequality

A natural place to begin to analyse the observed changes in the wage structure is to consider a model of changes in supply and demand. We then need to incorporate institutional features (such as minimum wages and trade unions) into the model.

Sources of Skill Premia: Supply and Demand

Rising wage inequality has been accompanied by an increase in the relative demand for skilled or educated workers. This is evident since, despite the increase in the relative supply of more skilled workers in many countries, their relative wage has also gone up, suggesting that relative demand for skilled workers has been rising faster than relative supply. In Table 2, for example, the proportion of graduates grew from 20.8 per cent of the population in 1980 to 34.2 per cent in 2004 in the United States. (In the United States the graduate measure is having a bachelor's degree or higher – that is, excluding people with some college who do not get a degree.) The equivalent figures from the United Kingdom were even more dramatic – the growth in graduates was from 5 per cent to 21 per cent over the same time period. However, at the same time the relative wages of graduates compared with those of non-graduates increased. In a competitive model of the labour market with skilled and unskilled workers, these facts can be reconciled by an increase in the relative demand for skilled workers.

A simple way to formalize this, following Katz and Murphy (1992), is in the context of a constant elasticity of substitution (CES) production function with two labour inputs:

$$Q_t = \left[\alpha_t (a_t N_s)_t^{\rho} + (1-\alpha_t)(b_t N_u)_t^{\rho}\right]^{1/\rho} \quad (1)$$

In Eq. (1) aggregate output is Q and is produced with college educated-equivalent skilled labour (N_s) and high school-educated equivalent unskilled labour (N_u) in period t. The parameters a and b represent skilled and unskilled augmenting technical change, α indexes the share of work activities of skilled labour and ρ is a parameter that determines the elasticity of substitution between skilled and unskilled labour $\left(\sigma = \frac{1}{1-\rho}\right)$. Skill-biased technological changes involve increases in a/b or α.

Assuming college and high school equivalents are paid their marginal product we can use Eq. (1) to solve for the ratio of marginal products of the two types of labour, W_s/W_u, and relative supplies of labour, N_s/N_u, in year t as:

$$\ln(W_s/W_u)_t = (1/\sigma)\left[D_t - \ln(N_s/N_u)_t\right] \quad (2)$$

where

$$D_t = \sigma\left[\ln(\alpha_t/(1-\alpha_t)) + \rho \ln(a/b)_t\right] \quad (3)$$

is a relative demand index of shifts favouring college equivalents and is measured in log quantity units. The impact of changes in relative skill supplies (N_s/N_u) depends on the elasticity of substitution, σ. The larger this parameter is, the bigger will be the effects of supply changes on relative wages. Eq. (3) shows that changes in D can arise from (disembodied) skill-biased technical change, non-neutral changes in relative prices or quantities of non-labour inputs and shifts in product demand.

Katz and Murphy (1992) implemented an empirical version of Eq. (2), replacing D with a linear time trend ('*trend*') for US data between 1963 and 1987. They estimate:

$$\ln(W_s/W_u)_t = \gamma_0 + \gamma_1 trend + \gamma_2 \ln(N_s/N_u)_t + v_t \quad (4)$$

finding $\hat{\gamma}_2$ to be significantly negative (equal to -0.709), implying an elasticity of substitution of about 1.4 ($\sigma = -1/\hat{\gamma}_2 = 1{:}41$), with a significant trend increase in the college premium of 3.3 per cent per annum ($\hat{\gamma}_1 = :033$). In the literature that has followed the estimates of the elasticity of substitution are typically in the 1.4–1.6 range (see, for example, the study of Autor et al. 2005). The main point to take away from these estimates is that there appears to be a systematic demand shift towards more skilled workers throughout the four last decades of the 20th century.

This is not to suggest that supply side changes are unimportant. Deviations of relative skill supplies from the trend are negatively associated with deviations of the relative wage from trend as suggested by $\gamma_2 < 0$. The *slowdown* of the growth of education in more recent cohorts is certainly one factor accounting for the increase in inequality as shown by Card and Lemieux (2001). But the most important factor over the longer run in accounting for the growth in educational wage differentials appears to be the trend demand shift towards the more skilled. The critical question then becomes: what could account for this change?

The Cause of Relative Demand Shifts: Technology or Trade?

To date, the two main explanations for the demand shift towards the more skilled are skill-biased technological change (SBTC) and increased international trade. We examine each of these in turn.

Skill-Biased Technological Change

Equation (3) above directly relates the change in the skill premia to SBTC. The idea is that new technologies such as information and communication technologies (ICTs) are complementary with the skills of more-educated workers. More-educated workers may find it easier to cope with the uncertainty surrounding new technologies in general, or may have a particular advantage in using ICT effectively. Rapid falls in the quality-adjusted prices of ICT or a more rapid investment in new technologies (for example, from higher R&D intensities) could therefore have shifted demand towards more-skilled workers.

There is now abundant empirical evidence that suggests that SBTC *is* an important and

international phenomenon (for example, see the survey in Bond and Van Reenen 2007). A typical analysis estimates the following cost share equation (usually for industries or workplaces):

$$\Delta SHARE = \beta_1 TECH + \beta_2 \Delta \ln(K/Y) + \beta_3 \ln(W_s/W_u) + e \quad (5)$$

where *SHARE* is the wage bill share of skilled workers, *TECH* is a measure of technical change, K is the capital stock, Y is value added, W_s/W_u is relative wages, Δ the difference operator and e an error term. This relationship can be derived from the stochastic form of a translog short-run variable cost function with labour as the two variables and physical capital and technological capital as the two fixed factors (for example, Berman et al. 1994). The test of skill-biased technical change is whether $\beta_1 > 0$, and the overwhelming preponderance of econometric evidence supports this finding.

An example of the genre is Machin and Van Reenen (1998), who examine this relationship using manufacturing data across many industries in seven OECD countries (the United States, the United Kingdom, France, Japan, Germany, Denmark and Sweden) in the 1970s and 1980s. In all of the countries examined they found that demand was shifting more quickly towards skilled workers in the more technologically advanced industries (that is, $\beta_1 > 0$ in Eq. (5)). This was robust to using either occupation or education as a measure of skills, using either R&D intensity or computer use as a measure of technology, and instrumenting own R&D with frontier (US) R&D. In most countries they also found evidence of capital–skill complementarity ($\beta_2 > 0$). Estimating versions of Eq. (5) in other countries, in non-manufacturing sectors (for example, Autor et al. 1998) and on more disaggregated plant-level data (for example, Doms et al. 1997) also appears to uncover evidence of SBTC.

There are several other sources of evidence on SBTC. Berman et al. (1998) report evidence of faster skill demand shifts occurring in the same sorts of industries in different countries, and one may view this as informing the SBTC hypothesis (to the extent that similar industries in different countries utilize similar technologies). A less-used alternative to test for SBTC is to regress the adoption of technologies on skill prices (that is, when skilled workers' wages rise relative to those of unskilled workers this should depress the incentive to adopt new technologies) or skilled labour supply – some evidence for this method is in Caroli and Van Reenen (2001) and Doms and Lewis (2006), and also supports SBTC. A third method is to directly estimate the production function or the cost function underlying the factor demand Eq. (5). This has also tended to uncover evidence of skill–technology complementarity (for example, Bresnahan et al. 2002). Finally, some authors have directly regressed individual wages on computer use or controlling for other factors (for example, Krueger 1993). In our view, this is a rather unsatisfactory test of SBTC, however, as computers are likely to be allocated to more productive workers, as has been found by several studies (Chennells and Van Reenen 1997; DiNardo and Pischke 1997).

Although we have stated the SBTC hypothesis in quite a blunt fashion, the influence of technical change almost certainly acts in a more subtle ways to affect outcomes as detailed case studies suggest (Blanchard 2004). For example, some econometric studies suggest that technical change operates through organizational changes (for example, through decentralizing or delayering hierarchies) that are typically associated with increased demand for skilled workers (Caroli and Van Reenen 2001; Bresnahan et al. 2002). Moreover, computerization does not simply involve increasing all skill demand, but it substitutes for different types of tasks. Autor et al. (2003) offer a more nuanced version of the SBTC hypothesis, arguing that computerization reduces the demand for routine tasks (for manual and non-manual workers) but results in an increase in demand for analytic or non-routine skills. Thus, routine non-manual tasks (for example, clerical work) may be replaced by computers, whilst some non-routine tasks done by manual workers (like cleaning) are largely unaffected by IT. The evidence on polarization of work referred to above where the 'middle' of

the wage distribution has suffered at the expense of the bottom as well as the top is in line with this. Building on upon these empirical observations, Autor et al. (2006) develop a model where IT replaces routine tasks to rationalize the experience of the 1990s when polarization of jobs occurred in the United States.

Overall then, there is strong support for the importance of SBTC. Some critics (most strongly expressed in Card and DiNardo 2002) argue that SBTC cannot be the reason for increased inequality because technical change is continuous whereas the change in wage inequality is episodic. Regardless of whether one agrees with the characterization of technical change, this misses the point that SBTC is meant to account for the longer-run pressure to increase skill demand (the D in Eq. (2)) and not necessarily the 'twist' in the wage structure in the 1980s. Similarly, the fact that inequality growth slowed down post-1995 whereas productivity growth accelerated does not disprove the SBTC argument, as the speed of technical change is not the same as the bias of technical change.

Increased International Trade

At first glance, the simple Heckscher–Ohlin model of trade offers a seemingly cogent explanation of why unskilled workers have faired badly in recent decades. Less-developed countries such as China and India have become integrated into the global economy as trade barriers and transportation and communication costs have fallen. Unskilled workers in the OECD counties now have to compete not only with workers at home but also with a large number of workers overseas. The influx of cheap goods produced with low-skill labour puts downward pressure on the wages and employment opportunities of unskilled workers in the West, and is responsible for the observed shifts in relative labour demand.

To model this we explicitly consider two regions: 'North,' which is skill-abundant and 'South' which is unskilled-abundant. There are four industries: tradable high-skill intensive, tradable low-skill intensive, non-tradable high-skill intensive and non-tradable low-skill intensive. The Stolper–Samuelson theorem establishes that relative wages in each country will depend on relative output prices of the tradable industries: the higher the relative price of the skill-intensive good, the higher the relative wage of the skilled workers. What happens when a small open economy in the North moves from autarky to free trade? The removal of trade barriers increases the relative price of the skill-intensive good and this means the skill premium rises in the North.

Although this model is coherent, it also offers several other predictions which turn out to be at odds with the data (see Desjonqueres et al. 1999, for extensive discussion of these predictions). First, the increasing specialization of the North in skill-intensive goods under free trade means that employment should shift *between* industries to skill-intensive industries. But because relative skill prices have risen we should expect to see that employment *within* industries shifts towards (the cheaper) unskilled workers. Decompositions of the increase in the aggregate employment share of skilled workers, however, almost all show that within industries there has been a strong shift towards skilled workers. This might be because the level of aggregation of industries is too high, but more disaggregated industries and even firm-level studies suggest that a sizable proportion is 'within'. Even more convincingly, Desjonqueres et al. (1999) show that non-traded sectors – such as hotels and wholesale outlets – also show a shift towards skilled workers (and an increase in the educational wage premium). This pattern of within-industry shifts is consistent with general SBTC, but inconsistent with the basic trade theory.

Second, we should observe that relative prices of the unskilled-intensive sectors should fall rapidly in the North. There is some evidence for this in the United States but there is no significant relationship for any other country (at least until the mid-1990s). Even in the United States the evidence from Krueger (1996) suggests that this relationship was only apparent after 1989, when wage inequality grew slowly. Finally, naive regressions that include import penetration and other trade variables in Eq. (5) generally find no role for these trade variables (for example, Machin and Van Reenen 1998). This does not take into

account the general equilibrium effects underlying the Heckscher–Ohlin model, of course.

Overall there is little support for the trade-based explanation of demand shifts. There are two caveats to this conclusion. First, most of these studies were based on data prior to the early 1990s when China started to become more of a major exporter. Second, trade might induce some of the skill biased technological change discussed in the previous section as suggested by Acemoglu (2002).

Labour Market Institutions

Research trying to reconcile cross-country differences in change in wage inequality has emphasized the role of labour market institutions that affect wages differently in different places. There are several features of this work, ranging from studies that look in detail across countries to those that focus on the role played by particular labour market institutions like minimum wages or trade unions.

Cross-Country Evidence

As discussed in Section "What Has Happened to the Wage Distribution?", there has been considerable heterogeneity in the evolution of relative wages across OECD countries since the 1970s. The rise in inequality was much stronger in the Anglo–Saxon countries (for example, the United States and the United Kingdom) than elsewhere (for example, France, Germany and Japan). Although the technology and/or trade shocks discussed in the previous subsections should be global events, the Continental European and Japanese economies have experienced a much greater increase in unemployment than the United States since the late 1970s. One view is that European unemployment and American inequality are 'two sides of the same coin' – institutional rigidities (and perhaps generous welfare benefits) placed a floor under the wages of unskilled workers in Continental Europe, resulting in increased unemployment rather than greater inequality. (There is a also a new, growing body of work arguing that tastes and social norms are important for explaining cross-country patterns of change; see, amongst others, Bénabou and Tirole 2006.)

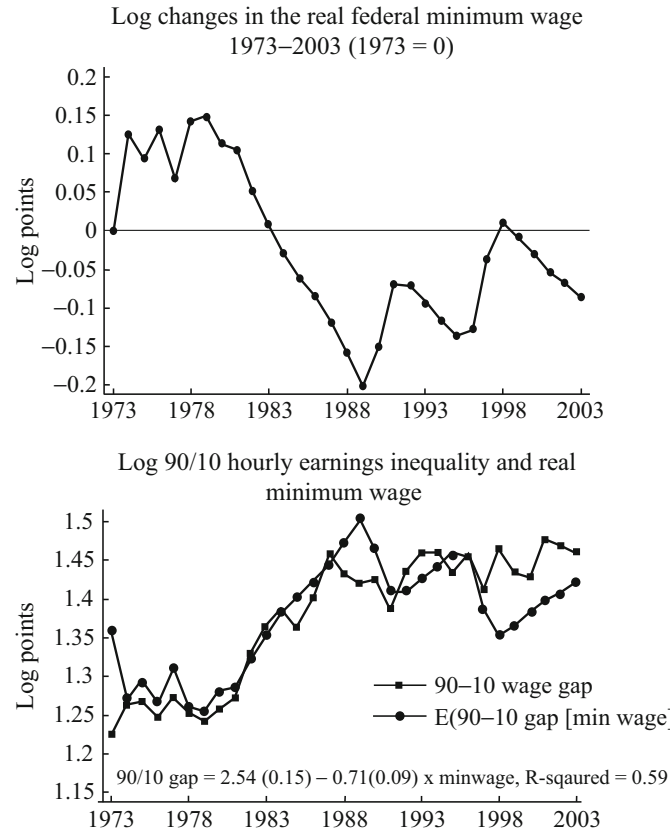
This is probably too crude. Nickell and Bell (1995) have shown that relative unemployment rates between skilled and unskilled workers did not rise by as much as would be expected in this simple model. Similarly, the cross-country correlation between the growth in unemployment and earnings inequality is not very strong (for example, Burniaux et al. 2006). Finally, European countries may have been better at keeping up the growth of supply of the quantity and quality of skills than in the United States and United Kingdom (although Table 2 shows that skill expansion in the United Kingdom was very rapid).

At the very least, the fact that wage inequality has not risen in the countries where minimum wages and/or union power remained strong suggests that institutions do have an important role to play.

Minimum Wages

There is much evidence that minimum wages compress wage differentials (DiNardo et al. 1996). In the United States the real value of the Federal minimum wage fell significantly during the 1980s, and some authors argue that this can account for all of the change in wage inequality (for example, Lee 1999). By the same token the uprating of the minimum wage in the 1990s helps explain the slowdown in wage inequality. As Card and DiNardo (2002) emphasize, the time series pattern is very strong – see Fig. 2.

A problem with the 'purely institutional' argument, however, is that it seems highly unlikely that the minimum wage can explain what is happening in the top half of the wage distribution. Analysis of the minimum wage suggests that the impact on workers above median wages is close to zero. Nevertheless, the most striking finding of the analysis in Section "What Has Happened to the Wage Distribution?" was that there appeared to be a secular increase in the 90–50 wage ratio since the late 1970s in the United States (and the United Kingdom). It is hard to reconcile these facts with the minimum wage-explains-all story. Similarly, when Autor et al. (2005) add the minimum wage to Eq. (4), although it has the expected negative sign it does little to reduce the long-run unexplained relative demand shift towards higher education wage differentials.



Wage Inequality, Changes In, Fig. 2 The time series relationship between the US federal minimum wage and wage inequality (*Source*: Autor et al. (2005))

Where the institutional story does better is in accounting for the dramatic increase in residual wage inequality in the bottom half of the wage distribution in the 1980s. This residual wage change was more episodic, and most of the change is plausibly accounted for by the minimum wage (and compositional effects – see below).

Another problem with the pure minimum wage explanation is that wage floors changed much less in other countries where wage inequality also rose. For example in the United Kingdom, the minimum wage system that operated at the time when wage inequality rose (the 'Wage Councils') only covered a relatively small proportion of the workforce (around 12 per cent at the time of abolition in 1993). Furthermore, during the 1993–9 time period when all non-agricultural minimum wages were abolished in the United Kingdom, wage inequality at the lower end actually started to stabilize (Dickens et al. 1999; Machin and Manning 1994).

Trade Unions and Imperfect Competition

As with minimum wages there is robust evidence that unions act to compress wage differentials (for example, Freeman 1980; Card 1996). Since unions have declined in the United States and the United Kingdom, this may be another institutional mechanism putting upwards pressure on wage inequality. Unionization rates fell from 25 per cent to 15 per cent between 1979 and 1998 in the United States, and from 53 per cent to 31 per cent in the United Kingdom over the same period. Gosling and Lemieux (2004) argue that union decline can account for over a third of the increase in male wage inequality in both countries over the 1983–98 period.

As with the minimum wage explanation, it is rather difficult to evaluate these statistical decompositions as they are not based on an underlying economic model. But it does seem rather implausible that unions could be the major explanation in the United States for the ongoing increase in the 90–50 ratio since (*a*) they comprise such a small part of the workforce and (*b*) their membership is mainly drawn from the bottom half of the wage distribution.

An alternative set of theories has emerged that emphasizes rents derived from imperfect competition (albeit from a different source from unions). This approach has frictions in the labour market that generate heterogeneous wages even for identical workers. Some more productive or technologically advanced firms may share quasi-rents to workers who are matched with them (for example, Van Reenen 1996). If the dispersion of these wage premia has increased over time, this could lie behind the increased wage inequality. For example, in Caselli (1999) firms experiment with the uncertain new technology, and some of those that are successful obtain higher productivity, resulting in higher wages for the workers with whom they are matched. To date, there is little hard empirical evidence on these theories, although Faggio et al. (2006) offer some evidence that firm productivity heterogeneity has increased and this is linked to firm wage inequality as Caselli's model would suggest.

Conclusions

There has been a dramatic increase in wage inequality since the late 1970s in the United States, the United Kingdom and other anglophone countries. A significant part of this is due to the growth of wage differentials between educational groups. We have argued that the fundamental reason for this is a long-run growth in the relative demand for skills driven by technology change (rather than trade). Changes in skill supply and institutional changes have affected the timing of how skill-biased technical change impacts upon the wage structure. The increase in inequality in the United States and the United Kingdom slowed down after 1990, but has continued to grow in the upper tail of the wage distribution, and wage inequality has started to rise in places previously characterized by stable wage structures (like Germany), indicating that explaining changing patterns of wage inequality remains high on the research agenda of empirical economists.

See Also

- ▶ Inequality (International Evidence)
- ▶ Minimum Wages
- ▶ Skill-Biased Technical Change
- ▶ Technology
- ▶ Trade and Poverty

Acknowledgment The Economic and Social Research Council has given financial support for this research through the Centre for Economic Performance. The usual disclaimer applies.

Bibliography

Acemoglu, D. 2002. Technical change, inequality and the labour market. *Journal of Economic Literature* 40: 7–72.

Autor, D., L.Katz, and M.Kearney. 2005. Trends in U.S. wage inequality: Re-assessing the revisionists. Working Paper No. 11627. Cambridge, MA: NBER.

Autor, D., L. Katz, and M. Kearney. 2006. The polarization of the U.S. labor market. *American Economic Review* 96: 189–194.

Autor, D., L. Katz, and A. Krueger. 1998. Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics* 113: 1169–1214.

Autor, D., F. Levy, and R. Murnane. 2003. The skill content of recent technological change: An empirical investigation. *Quarterly Journal of Economics* 118: 1279–1333.

Bénabou, R., and J. Tirole. 2006. Belief in a just world and redistributive politics. *Quarterly Journal of Economics* 121: 699–746.

Berman, E., J. Bound, and Z. Griliches. 1994. Changes in the demand for skilled labor within U.S. manufacturing industries: Evidence from the annual survey of manufacturing. *Quarterly Journal of Economics* 109: 367–398.

Berman, E., J. Bound, and S. Machin. 1998. Implications of skill-biased technological change: International evidence. *Quarterly Journal of Economics* 113: 1245–1280.

Blanchard, O. 2004. The economic future of Europe. *Journal of Economic Perspectives* 18(4): 3–26.

Bond, S., and J. Van Reenen. 2007. Micro-econometric models of investment and employment. In *Handbook of econometrics*, ed. J. Heckman and E. Leamer, Vol. 6. Amsterdam: North Holland.

Bound, J., and G. Johnson. 1992. Changes in the structure of wages in the 1980s: An evaluation of alternative explanations. *American Economic Review* 83: 371–392.

Bresnahan, T., E. Brynjolfsson, and L. Hitt. 2002. Information technology, workplace organization and the demand for skilled labour: Firm-level evidence. *Quarterly Journal of Economics* 117: 339–376.

Burniaux, J., Padrini, F., and Brandt, N.. 2006. Labour market performance, income inequality and poverty in OECD countries. Working Paper No. 500, Economics Department, OECD.

Card, D. 1996. The effects of unions on the structure of wages. *Econometrica* 64: 957–979.

Card, D., and J. DiNardo. 2002. Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics* 20: 733–783.

Card, D., and T. Lemieux. 2001. Can falling supply explain the rising return to college for younger men? *Quarterly Journal of Economics* 116: 705–746.

Caroli, E., and J. Van Reenen. 2001. Skill-biased organizational change? Evidence from a panel of British and French establishments. *Quarterly Journal of Economics* 116: 1449–1492.

Caselli, F. 1999. Technological revolutions. *American Economic Review* 89: 78–102.

Chennells, L., and J. Van Reenen. 1997. Technical change and earnings in British establishments. *Economica* 64: 587–604.

Desjonqueres, T., S. Machin, and J. Van Reenen. 1999. Another nail in the coffin? Or can the trade based explanation of changing skill structures be resurrected? *Scandinavian Journal of Economics* 101: 533–554.

Dickens, R., S. Machin, and A. Manning. 1999. The effects of minimum wages on employment: Theory and evidence from Britain. *Journal of Labor Economics* 17: 1–22.

DiNardo, J., N. Fortin, and T. Lemieux. 1996. Labor market institutions, and the distribution of wages, 1973–1992: A semi-parametric approach. *Econometrica* 64: 1001–1044.

DiNardo, J., and S. Pischke. 1997. The returns to computer use revisited: Did pencils change the wage structure too? *Quarterly Journal of Economics* 112: 291–303.

Doms, M., and E.Lewis. 2006. Labor supply and personal computer adoption, Research Paper No. 06–10, Federal Reserve Bank of Philadelphia.

Doms, M., T. Dunne, and K. Troske. 1997. Workers, wages and technology. *Quarterly Journal of Economics* 112: 253–289.

Faggio, G., J. Van Reenen, and K. Salvanes. 2006. Increasing inequality of firm productivity. In *Mimeo, centre for economic performance*. London: School of Economics.

Freeman, R. 1980. Unionism and the dispersion of wages. *Industrial and Labor Relations Review* 34: 3–23.

Goldin, C., and L.Katz. 1999. The returns to skill in the United States across the twentieth century. Working Paper No. 7126. Cambridge, MA: NBER.

Goldin, C., and L. Katz. 2001. Decreasing (and then increasing) inequality in America: A tale of two half centuries. In *The causes and consequences of increasing inequality*, ed. F. Welch. Chicago: University of Chicago Press.

Goos, M., and A. Manning. 2007. Lousy and lovely jobs: The rising polarization of work in Britain. *The Review of Economics and Statistics* 89: 118–133.

Gosling, A., and T. Lemieux. 2004. Labor market reforms and changes in wage inequality. In *Seeking a premier league economy*, ed. D. Card, R. Blundell, and R. Freeman. Chicago: Chicago University Press.

Juhn, C., K. Murphy, and B. Pierce. 1993. Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101: 410–442.

Katz, L., and D. Autor. 1999. Changes in the wage structure and earnings inequality. In *Handbook of labor economics*, ed. O. Ashenfelter and D. Card, Vol. 3. Amsterdam: North-Holland.

Katz, L., and K. Murphy. 1992. Changes in relative wages, 1963–87: Supply and demand factors. *Quarterly Journal of Economics* 107: 35–78.

Krueger, A. 1993. How computers have changed the wage structure: Evidence from microdata, 1984–1989. *Quarterly Journal of Economics* 108: 33–60.

Krueger, A. 1996. Skill intensity and industrial price growth. *Industrial Relations Research Association Proceedings* 48: 11–18.

Lee, D. 1999. Wage inequality in the United States during the 1980s: Rising inequality or falling minimum wages. *Quarterly Journal of Economics* 114: 997–1023.

Lemieux, T. 2006. Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review* 96: 461–498.

Machin, S. 1996. Wage inequality in the UK. *Oxford Review of Economic Policy* 7: 49–62.

Machin, S., and A. Manning. 1994. Minimum wages, wage dispersion and employment: Evidence from the U.K. wages councils. *Industrial and Labor Relations Review* 47: 319–329.

Machin, S., and J. Van Reenen. 1998. Technology and changes in skill structure: Evidence from seven OECD countries. *Quarterly Journal of Economics* 113: 1215–1244.

Machin, S., and A. Vignoles. 2005. *What's the good of education? the economics of education in the UK*. Princeton: Princeton University Press.

Nickell, S., and B. Bell. 1995. The collapse in demand for the unskilled and unemployment across the OECD. *Oxford Review of Economic Policy* 11: 40–62.

Spitz-Oener, A. 2006. Technical change, job tasks and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics* 24: 235–270.

Van Reenen, J. 1996. The creation and capture of economic rents: Wages and innovation in a panel of UK companies. *Quarterly Journal of Economics* 111: 195–226.

Wages Fund

John Vint

Abstract

The wages fund doctrine was an important element in the classical analysis of the labour market – elements of the wages fund doctrine are to be found in the *Wealth of Nations* (1776) – and articles attempting to defend it were still being produced over a hundred years later. Its longevity was due to its success in generating a wide range of economic predictions. It is noted for John Stuart Mill's recantation in 1869 – a rare case of an important doctrine being explicitly rejected by a major political economist.

Keywords

Avances; Malthus, T. R; Marcet, J; Marginal productivity theory; Martineau, H; McCullough, J. R; Mill, J. S; Sidgwick, H; Smith, A; Strikes; Wages fund; Walker, F. A

JEL Classifications

B1

The wages fund doctrine was an important element in the classical analysis of the labour market – elements of the wages fund doctrine are to be found in the *Wealth of Nations* in 1776 – and articles attempting to defend it were still being produced over a hundred years later. The wages fund doctrine began to take shape in the work of the early classical writers and became more rigid later. On the one hand, it was popularized by Marcet (1816) and Martineau (1832) among others, and employed in more vulgar forms in political debates; on the other, it was used much more carefully and technically by the classical economists to produce a wide range of economic predictions. It is noted for John Stuart Mill's recantation in 1869 – a rare case of an important doctrine being explicitly rejected by a major political economist.

The approach to the wages fund doctrine adopted by the classical writers varied over the period but by the middle of the 19th century there was a generally accepted analysis, from which Mill dissented. In this article I will begin by examining this recantation view of the doctrine and then outline the various phases in its development and decline.

In his book *On Labour,* which appeared in 1869, W.T. Thornton presented a critique of supply and demand analysis and the wages fund doctrine. John Stuart Mill reviewed the book in the *Fortnightly Review* for May and June 1869, and in this review he made his famous recantation from the wages fund doctrine. Mill sketched what he understood the accepted doctrine to be:

> There is supposed to be, at any given instant, a sum of wealth, which is unconditionally devoted to the payment of wages of labour. This sum is not regarded as unalterable, for it is augmented by saving, and increases with the progress of wealth; but it is reasoned upon as at any given moment a predetermined amount. More than that amount it is assumed that the wages- receiving class cannot possibly divide among them; that amount, and no less, they cannot but obtain. So that, the sum to be divided being fixed, the wages of each depend solely on the divisor, the number of participants. In this doctrine it is by implication affirmed, that the demand for labour not only increases with cheapness, but increases in exact proportion to it, the same aggregate sum being paid for labour whatever its price may be. (1869, p. 515)

The predetermined fund is usually conceived of as a stock of wage goods or necessaries, and if this stock is changed from one production period to the next, while the labour force remains constant, the real wage rate would also change. Some classical economists discussed the consequences of a change in the money fund under circumstances where the fund of wage goods remained the same. One line of argument was that the workers could consume only wage goods and not luxuries, and that therefore the increased expenditure on the fixed stock of goods would lead to an increase in prices, leaving the real wage rate unchanged. This can be regarded as

the 'rigid' version of the theory. The fact that not all classical writers subscribed to this version all the time was ultimately an important feature in its demise.

The origins of the doctrine in British political economy are to be found in the work of Adam Smith, although he was heavily influenced by the Physiocrats' notion of *avances*. In the *Wealth of Nations* Smith clearly proposed the notion that wages are advanced to workers by capitalists from capital, although there is nothing to suggest that these advances are predetermined or fixed. Moreover, while the wages fund theory presents a competitive solution to the determination of the wage rate, Smith argued for an uncompetitive solution 'upon all ordinary occasions' (1776, p. 82) due to the superior bargaining position of the employers.

In the first edition of the *Essay on the Principles of Population* 1798, Malthus put forward all the elements of the wages fund doctrine, including the argument that increased money payments would be offset by increased prices, leaving the real wage rate unchanged. However, in the first edition of the Malthus's *Principles of Political Economy* (1820), the wages fund was not well developed. Labour demand was loosely related to growth in capital or resources. By contrast, in the second edition of the *Principles of Political Economy* (1836), Malthus's discussion of wages is much fuller and much closer to a wages fund approach.

There is no simple, consistent statement of the wages fund doctrine as outlined above to be found in Ricardo's *Principles of Political Economy* (1817), or any of his other published works or correspondence. In the chapter on machinery, Ricardo does present an example which makes use of the logic of the wages fund, but there are other passages in the *Principles* and in his correspondence where Ricardo's approach to wages runs directly counter to the doctrine. For example, in correspondence with Trower, Ricardo (1951, VIII, p. 258) argued that under certain circumstances workers can consume luxuries.

The mature phase, when the accepted version fully emerged, is associated with the work of Marcet and McCulloch, and the doctrine was applied to a range of economic issues by Torrens, Senior and J.S. Mill. McCulloch has been regarded by prominent historians of economic thought such as Bonar (1885, p. 272) and Cannan (1893, p. 263) as the author of the analysis which had originated in Marcet's *Conversations on Political Economy* (1816), although it had, in fact, been originally expounded by Malthus in 1798 (see Vint 1994, pp. 77–88).

McCulloch argues in the *Principles of Political Economy* (1825) that a country's ability to employ workers depended on the existence of an 'amount of the accumulated produce of previous labour'. The wage rate depended on the 'proportion which the whole capital bears to the whole amount of the labouring population' (1825, p.173). He argued that wages do not depend on the amount of money allocated to labourers – if the mount of money halved but the quantity of wage goods remained the same the labourer 'would carry a smaller quantity of pieces of gold and silver to market than formerly; but he would obtain the same quantity of commodities in exchange for them (1825, p. 174). This was the rigid fund of Malthus.

It was the Malthus–McCulloch rigid version of the wages fund doctrine which underpinned some of the later popularizations used to argue against the efficacy of strikes and the role of trades unions. But as Taussig pointed out (1896, pp. 239–40), the classical economists themselves, up until Mill's *Principles* in 1848, did not use the doctrine in this way. McCulloch himself, for example, eschewing the wages fund doctrine in this context, made a powerful case *for* trades unions in 1824; Torrens (1834) made use of the wages fund analysis to show that unions could act to raise wages. Mill did use the wages fund doctrine to deny the effectiveness of union action in 1848 but his case was heavily qualified.

Among the popularisers, Harriet Martineau made use of the rigid fund in 'Manchester Strike', one of the tales from the *Illustrations of Political Economy* (1832), to argue that strikes were futile; William Ellis argued that in a combination to raise wages 'success is impossible' because 'the capital

out of which the increased wages are to be drawn does not exist' (1854, p. 224).

There were clearly potential weaknesses in the rigid version of the doctrine – were wages actually paid from capital; was the fund predetermined; could workers consume only wage goods? Despite these potential flaws the doctrine lasted for almost a century, and from the mid-1820s was an accepted part of classical theory. One potential explanation for its longevity is that it was used successfully by the major political economists such as Senior (1830) and J.S. Mill to produce a wide range of predictions relating, for example, to the effects of the introduction of machinery on the wage rate, the impact of various kinds of war loans on wages and the effects of landlord absenteeism (see Vint 1994, pp. 124–75).

Mill's recantation came at the end of a decade of change for the British trade union movement. Events on the ground were important – there was an intense political debate concerning the legal status and role of trades unions, culminating in the appointment of Royal Commission of Inquiry in 1867. Alongside these events there was an important theoretical discussion concerning the wages fund doctrine in the work of Fawcett (1860), Longe (1866) and most importantly Thornton (1869). Longe argued that wages were not paid from capital, and both he and Thornton argued that the wages fund was not predetermined. In the 'recantation' Mill attacked the theory, arguing that the demand for labour does *not* increase with cheapness, that the wages fund is not predetermined and that the money funds in the hands of employers were flexible and could be bargained for by workers. He then went on to produce his famous recantation statement:

> The doctrine hitherto taught by all or most economists [including myself], which denied it to be possible that trade combinations can raise wages, or which limited their operation in that respect to the somewhat earlier attainment of a rise which the competition of the market would have produced without them, – this doctrine is deprived of its scientific foundation, and must be thrown aside. The right and wrong of the proceedings of Trades' Unions becomes a common question of prudence and social duty, not one which is peremptorily decided by unbending necessities of political economy. (1869, pp. 517–18)

Thus in a few short paragraphs Mill apparently disposed of a central tenet of classical economics and one which had a long history. There were attempts to revive the doctrine by Cairnes (1874) and others, but these failed. Cairnes's arguments were thoroughly demolished by F.A. Walker (1875) in the first comprehensive review of the debate. In 1879 Sidgwick also reconsidered the controversy in the *Fortnightly Review.* After agreeing that 'Professor Walker's argument gives a *coup de grâce* to the old wages-fund theory', Sidgwick threw out a challenge to Walker and other economists to put something in its place (1879, p. 411). This challenge and the response of Walker and others to it generated what Gordon referred to as the 'second round debate' (1973, pp. 23–31) and which resulted in the eventual development of marginal productivity theory. Until Sidgwick's watershed contribution, the point of reference for the discussion of wage theory was the debate which began in the 1860s, but thereafter subsequent writers made fewer references to the earlier controversy. There continued to be defences of the doctrine, but often the focal point was Walker's work, and none of these efforts succeeded in undoing the damage done earlier.

See Also

- ▶ Cairnes, John Elliott (1823–1875)
- ▶ Machinery Question
- ▶ McCulloch, John Ramsay (1789–1864)
- ▶ Mill, John Stuart (1806–1873)
- ▶ Senior, Nassau William (1790–1864)
- ▶ Strikes
- ▶ Walker, Francis Amasa (1840–1897)

Bibliography

Bonar, J. 1885. *Malthus and his work*. New York: Allen and Unwin.

Cairnes, J.E. 1874. *Some leading principles of political economy newly expounded*. London: Macmillan.

Cannan, E. 1893. *A history of the theories of production and distribution*. London: Percival.
Ellis, W. 1854. In *Lessons on the phenomena of industrial life, and the conditions of industrial success,* ed. Rev. R. Dawes. London: Groombridge.
Fawcett, H. 1860. Strikes: Their tendencies and remedies. *Westminster Review* NS 18: 1–23.
Gordon, S. 1973. The wages fund controversy: The second round. *Quarterly Journal of Economics* 93: 501–522.
Longe, F.D. 1866. *A refutation of the wages fund theory of modern political economy as enunciated by Mr. Mill and Mr. Fawcett, M.P.* London: Longmans, Green.
Malthus, T.R. 1798. *An essay on the principles of population as it affects the future improvement of society, with remarks on the speculations of Mr. Godwin, M. Condorcet, and Other Writers*. Repr. by Royal Economic Society as *First Essay on Population.* London: J. Johnson, 1926.
Malthus, T.R. 1820/1836. *Principles of political economy,* Variorum edition, ed. J. Pullen. London: Cambridge University Press for the Royal Economic Society, 1989.
Marcet, J. 1816. *Conversations on political economy, in which the elements of that science are familiarly explained*. London: Longman, Hurst, Rees, Orme and Brown.
Martineau, H. 1832. The Manchester strike. In *Illustrations of political economy,* No. 7. London: Charles Fox.
McCulloch, J.R. 1825. *Principles of political economy.* London: Ward, Lock and Co.
Mill, J.S. 1848. *Principles of political economy,* vols. 2 and 3 of *Collected works of John Stuart Mill,* ed. J.M. Robson. Toronto: University of Toronto Press, 1965.
Mill, J.S. 1869. Thornton on labour and its claims. *Fortnightly Review* 505–18: 680–700.
Ricardo, D. 1817. Principles of political economy. In *The works and correspondence of David Ricardo*, vol. 1, ed. P. Sraffa. Cambridge: Cambridge University Press, 1951.
Ricardo, D. 1951. In *The works and correspondence of David Ricardo*, ed. P. Sraffa. Cambridge: Cambridge University Press, 1951.
Senior, N. 1830. *Three lectures on the rate of wages*. London: John Murray.
Sidgwick, H. 1879. The wages fund doctrine. *Fortnightly Review,* 401–413.
Smith, A. 1776. In *An inquiry into the nature and causes of the wealth of nations*, ed. R.H. Campbell and A.S. Skinner. Oxford: Oxford University Press, 1976.
Taussig, F.W. 1896. *Wages and capital*. London: Macmillan.
Thornton, W.T. 1869. *On labour: Its wrongful claims and rightful dues, its actual present and possible future*. London: Macmillan.
Torrens, R. 1834. *On wages and combinations*. London: Longman, Rees, Orme, Brown, Green and Longman.
Vint, J. 1994. *Capital and wages: A Lakatosian history of the wages fund doctrine*. London: Edward Elgar.
Walker, F.A. 1875. The wage-fund theory. *North American Review* 120(246): 84–119.

Wages in Classical Economics

Krishna Bharadwaj

Recent discussions stimulated by Piero Sraffa's editorial introduction and Commentaries on Ricardo (*Works*, Vols I–XI) and his slim but significant volume, *Production of Commodities by Means of Commodities* (1960) have brought to light the distinctive characteristics of the approach and structure of classical theory of value and distribution as contrasted with those of the marginalist theory which has dominated since the 1870s. The role of wages in classical theory needs to be perceived within the structure of its value and distribution theory, which analysed the central questions as to how surplus is generated, appropriated and distributed in a circular process of reproduction and further investigated into how these shaped and were, in turn, shaped by the process of accumulation. A central notion to the theory is that of surplus or 'social net product', defined as gross output of the economy produced during the chosen period minus 'productive consumption'; the latter, being the material means of production and the requisites for the sustenance of labour engaged in production. Through the various stages of theoretical advancement, from Petty and Boisguillebert to Smith, Ricardo and Marx, the various categories and constituents of net product, and of the means of production, the classes of surplus appropriators and sharers, the forms of exchange and rules of surplus distribution, altered, reflecting historical developments and the relevant analytical perceptions.

Wages were treated as a part of the productive consumption, an essential material necessity of production, whatever may be the historical form of labour (and correspondingly, the *form* of its revenue). The sustenance of the labour had to be provisioned as a prior condition: the Physiocrats were the first to theorize about wages as being 'advances' of subsistence. The necessity for wages to be advanced was particularly evident in

agriculture, the central focus of their inquiry. The idea of a 'uniform', 'given wage' materialized along with labour becoming progressively a 'commodity' and with the establishment of capitalist relations. The non-wage incomes were considered as residual incomes, as paid out of surplus emerging at the end of the productive cycle; hence not a 'material necessity' as was the sustenance of labour; wages were thus seen as physical costs of reproducing labour, in contrast to, as material incentives to induce efforts or disutility of effort as in later theory.

Adam Smith conceptualized features of capitalist production (reflected in the 'net product' no more confined solely to agriculture as in the Physiocrats, and in the emergence of the free wage labourer and of the 'capitalist'), of distribution (the tripartite division of social classes and the respective revenues – wages, profits and rents – with their diverse origins and nature), and of exchange (reflected in the relevant categories of 'natural price' and 'market price'); and, formalized the regime of competition (the tendency towards uniform rate of profit and wages). Ricardo, while accepting the general framework of analysis, criticized certain inconsistencies and ambiguities of Smith's, particularly in relation to the theory of distribution (profit); namely, Smith's suggestion that the rate of profit was determined by competition of capitals and that the natural rates of wages, profits and rents were determined independently of each other and were thus the independently determined 'causes' of 'natural price'. Ricardo focused on the question of distribution, purporting to demonstrate that 'profits depend upon wages'. It is in the determination of the rate of profit, defined as the social net product (or surplus, as above, after 'getting rid of rent'), divided by the value of capital ('productive consumption'), that Ricardo faced the need to have a consistent theory of value to measure the heterogeneous aggregates involved in this determination (see Sraffa's introduction to *The Collected Works of David Ricardo*, Vol. I, 1951). Ricardo formulated and used the labour theory of value for the purpose.

As brought out by Sraffa (1951), it is in Ricardo's *Essay on Profits* (1815) and his generalization of the corn rate of profit of the *Essay* into the general rate of profit in *Principles*, using the labour theory of value, that the underlying structure of the classical theory of value and distribution becomes transparently evident. In Marx, the same basic structure continues and his scheme of prices of production is presented in a many-commodities framework in Sraffa's *Production of Commodities* (1960). We find therein, given social output levels and composition, given methods of production and given the wage, under the competitive assumption of uniformity of profit and wage, the rate of profit and prices of products are simultaneously determined.

This structure implies that the determination of prices and distribution would need to be carried out, taking the output levels and methods of production and wage as provisional data; this is not to rule out the interaction among accumulation (changes in output), changes in technology and changes in wages – indeed, the classical writings, as we see below, were deeply concerned about these interactions. The analysis is carried out in separate stages so that, for the derivation of distributive shares, at any stage, these are taken as provisional data.

The determination of wages itself is explained in classical theory in terms of a variety of historical and socio-economic factors. The role of wages, as 'given' in the value-scheme, needs to be contrasted with the neoclassical theory wherein the prices of 'factors of production' along with prices of commodities are determined simultaneously with quantities of outputs, employing the same mechanism of demand and supply relations for factors and commodities; for which purpose, factor endowments, technological possibilities and consumer preferences are taken as data. Given these, the relevant supply and demand functions are generated on the basis of the universal application of the price-guided 'substitution' principle.

In the very early theories, the idea that subsistence wage was a physiological minimum appeared as an axiom derived from the prevalent conditions under which the ordinary, unskilled labour was performed. However, even with the Physiocrats, the idea of a 'given' wage (rather

than a *fixed* minimum subsistence) was emerging prominently. While a tendency for wages to be restored to a certain norm was recognized, the norm itself was determined by customs, conventions, 'political necessity' (as in Steuart) and variations in it were discussed in the context of the effect of accumulation, migration or changes in population or, vice versa.

Adam Smith, with his distinction between 'natural' and 'market' price of labour was to synthesize preceding discussions on wages into a much more interesting theory. We find in Smith, a graphic expression of the inequitous bargaining position of 'masters' and workmen and the ensuing social struggle. Wage is also influenced by the pace of accumulation so that the 'norms' (or 'natural price') of labour would be low, medium or high depending upon the stage of development of the country and while Smith acknowledged a tendency for wages to gravitate to the natural level, (so that, he argued, that money wages would move sympathetically with prices of provisions), the norm itself was variable, depending upon whether the economy was declining, stationary or progressive. Smith opposed, in fact, the Mercantilist belief (and policy prescription) that low wages were necessarily advantageous as they stimulated hard work and afforded a commercial advantage through cheapness of the products; Smith argued, on the contrary, that 'a plentiful subsistence' stimulates productivity ('where wages are high, accordingly, we shall always find the workmen more active, diligent and expeditious, than when they are low' p. 81); Smith highlighted the importance of stimulating productivity of labour as it was the surplus produce over and above the necessary wage that would be the fund for accumulation. It was the annual produce of productive labour which constituted the wealth of a nation and not the accumulation of species (Smith's emphasis on division of labour as stimulating productivity of labour arises from such a perspective). What is characteristic of the view, and distinctly different the marginalist theory, is that there is no necessary functional relation between labour productivity and the wage. He did not however define any quantitative limit on 'plentiful subsistence' and did not draw the implications of the inverse wage-profit relation which Ricardo focused upon, under, however, 'given methods of production'.

Smith discussed, following Cantillon, wage differentiation among different categories of workers arising due to different degrees of 'hardship and ingenuity'. These differences cannot be accurately measured but Smith referred to 'the higgling and bargaining' on the market from which arises a 'scale', which, once formed, shows little variation. The subsistence of an unskilled labour was taken as the basic-level norm, with wages of other categories maintaining a proportion to it. The levels of wage were not explained by 'productivity' of the worker (if they were, a circularity in reasoning would be involved) but were the cost of reproduction of labour of the particular kind at the *norm* of maintenance which was generated by custom, convention, cost of training and of rearing the requisite skills and, in general, by 'the higgling and bargaining of the market'. That wages were basically looked upon from the point of view of the needs of reproduction of labour is evident when, for example, Smith argues that the minimum wage or the natural price of labour is lower for the free workman than for the slave. The slave is 'dearer' than the free labourer because the latter looks after his 'wear and tear' whereas the 'fund for replacing "wear and tear" is wastefully and disorderly administered by a "negligent master or careless overseer"' (p. 81). It is also evident in the provision for cost of training in the wage of a skilled labourer.

In Ricardo, the distinction between 'natural wage' and 'market wage' is analysed more systematically: while the natural price of labour is considered to be 'that price which is necessary to enable the labourers, one with another, to subsist and perpetuate their race, without increase or diminution', it is not necessarily the physiologically fixed minimum level of wages. It can vary from region to region and time to time and incorporates a large element of 'custom and habit', which may render 'comforts' necessaries. There remains active, however, a tendency whereby fluctuations in wages (i.e. in 'market price of labour') gravitate to restore the natural wage. The forces that generate these deviations were analysed by Ricardo in

terms of 'supply' and 'demand' for labour; when the market price of labour exceeds the natural price, a stimulus is provided for an increase in population; consequently, increasing the supply of labour. The reverse is the case when the market price of labour falls short of the 'natural price'. Thus, given the demand for labour at any time, the adjustment occurred through the changes in the supply of labour – which adaptations, however, could be slow and may lead sometimes to a higher market price for a long period. The demand for labour would itself be dependent on the pace of accumulation and increase in capital. Ricardo distinguished between the two effects of accumulation: the 'increase' of capital (defined as 'food, clothing, tools, raw materials, machinery, necessaries of labour etc.') could affect the natural wage, if, along with the increase in the quantity of capital, its 'value' also increased. By the latter, Ricardo referred to the increasing difficulty of production (requiring higher labour) of the means of production, particularly food and necessaries of labour. Such a difficulty was envisaged by Ricardo especially in the production of food; in which case, the natural wage would increase along with the prices of provisions (unless the non-food necessaries of labour fell in value to compensate for the rise in food). If such is not the case, the natural price of labour would be unaffected. However, in either situation, Ricardo argued that there would be a tendency for the 'market price of labour' to rise, following the accentuation of demand for labour and this would set up the tendency for supply of labour to so adjust as to restore the wages to their natural level. How rapidly or closely this could happen therefore depended upon the effects of accumulation on the natural wage and the extent of stimulus given to demand for labour.

Certain clarifications need to be made regarding the peculiarities of Ricardo's analysis of the 'supply and demand factors'. First, while the 'supply' and 'demand' come into play to determine the variations of 'market price of labour', these are fluctuations around a 'natural wage' (and, anchored to it) which is determined exogeneously. For example, nowhere does Ricardo argue that the wage would be pushed permanently below the natural wage *in order to* achieve *full* absorption of labour. (The logical inference in the case of the marginalist supply and demand determined equilibrium would be that, at the 'equilibrium' wage, there should be full utilization of labour.) Smith, too, emphasized the conventional limits to depressing wages below a certain minimum, adopted by 'common humanity'. (In *Theory of Moral Sentiments*, he was to elucidate the rise of conventions and morality which renders a certain social order viable.) The second peculiarity is that Ricardo's population dynamics works in order to adjust the supply of labour to a demand that is generated by the process of accumulation which is independently determined. This is to be contrasted with the supply and demand mechanism that works in the 'factor-markets' in the neoclassical theory. In Ricardo's theory, it is the aggregate supply of labour that is altering. In the marginalist theory, the 'factor endowments' are given so that it is the relative demands and supplies of the different factors of production which are generated because of the price-guided substitution that occurs in the commodity and factor markets. The interplay of supply and demand forces in the labour market can no more be separated from other, factor and output, markets as the determination of wages in the labour market can happen only simultaneously with the determination of all other prices and all other quantities. In Ricardo, the demand for labour generated by accumulation (and, the variations in outputs that this process involves) is taken as independently given and the supply of labour adjusts only if market wages deviate from the 'natural'. Further, Ricardo does not posit any positive functional relationship between the growth of stock of capital and the rate of profit – as is presumed by 'modern' interpretations. While a certain minimum rate of profit was expected for investment to be positive, no monotonic positive relation such as above was envisaged, particularly for the determination of the rate of profit. Profits depended on wages alone. Further, an increase in capital may be accompanied by a reduction of demand for labour (as in the machinery question). Secondly, as noted above, the effect of accumulation on the natural wage

depends, according to Ricardo, on the conditions of production of the wage goods alone.

In the period after Ricardo, when his theory was already facing opposition, the Ricardians (James Mill and John Stuart Mill, in particular) appear to have gradually subordinated the notion of natural wage to the idea that wage is determined by the proportions of capital to labour. James Mill, in his *Elements*, discussed only the variations in wages caused by the disproportionality between population and capital, without any reference in the discussions to natural wage. The notion of wages fund (which began initially with the idea that the demand for labour was limited by the agricultural surplus available for 'advances') was developed as a proxy for capital, representing demand for labour and ultimately culminated into the wages fund doctrine which argued, on the basis of a *fixed* wage fund, that the only lasting means of improving the conditions of labour (i.e. a sustained rise in wages) was through constraints on the growth of the labouring poor. In his later years, J.S. Mill recanted his position.

Marx carried forward and extended in new directions the basic framework of value and distribution of Ricardo whose critical clarification had become necessary particularly since the later 'modifications' and 'extensions' by the Ricardians has obfuscated and corroded that structure. While continuing to hold allegiance to the idea of a given wage, Marx emphasized the influence of the historical elements in its determination. He savagely attacked the Malthusian population dynamics (which was raising its head in the wage fund doctrine) and instead focused on the historical process by which labour-power becomes a 'free' commodity, in the dual sense: freed of personal bondage and also freed of the means of sustenance and production. This situation of the capitalist, in complete control of the means of production and labour, possessing nothing but labour-power to sell, engenders the inequitous capital-labour relation, with the capitalist in command of the labour-process. The wage therefore is influenced by the state of class struggle. Marx displaces the importance of 'population adaptations' in the earlier theories by his concept of the 'reserve army of the unemployed'. He recognizes thereby that the primitive accumulation process that turns labour into a commodity generates a chronic but fluctuating pool of the unemployed and the size of this reserve army not only acts on the supply side of labour but the capitalist strategy of controlling the capital–labour relation is influenced by the size of the reserve army; it being a potential instrument for weakening the bargaining strength of the workers. Marx discusses the various strategies that the capitalists employ in order to strengthen the control over the labour process and maximize the difference between labour productivity and the wage paid out. The strategies discussed are in terms of mechanization, organizational systems, wage systems etc. Not only is the wage therefore not functionally linked to productivity but there are forces that precisely play on their differences. Marx, with his focus on changes in methods of production and on the process of accumulation probed more deeply on the interaction between these and wages. The process of capitalist accumulation was crisis-ridden and Marx visualized a number of contradictions arising in the system which could be potential causes of the ultimate breakdown. For example, the increasing relative immiserization of labour could create a realization problem. He also visualized the 'anarchy' of capitalist production leading to a crisis of disproportionalities. A long-term decline in the rate of profit could also threaten the system. These suggestions were neither worked out fully and rigorously, nor were their mutual compatibilities examined closely. This is partly unavoidable in a complex analysis of the interactions among distribution, accumulation and technology. These are bound to be influenced by a variety of forces and patterns of interdependence among variables specific to particular junctures. It would be precisely for this reason that the simple analytical core of the value and distribution theory needed to be firmly based in order to make sorties into the more complex analyses. Marx did demonstrate that, even with the more limited analytical tools at his disposal, he could venture a long distance. This was evident in his masterly analysis, in *Capital*, Vol. I, of the dynamics of the capital–labour relations and the evolution of wages and profits during the various stages of the growth of the capitalist system.

See Also

▶ British Classical Economics

Bibliography

Marx, K. 1890. *Capital*, 4th ed. London: Pelican, Marx Library. 1976.
Mill, J. 1821. *Elements of political economy*. London: Baldwin, Craddock & Joy.
Mill, J.S. 1848. *Principles of political economy*. In ed. W. Ashley. London: Longman, Green & Co., 1909.
Ricardo, D. 1821. Principles of political economy. In *The works and correspondence of David Ricardo*, ed. P. Sraffa and M.H. Dobb. Cambridge: Cambridge University Press. 1951.
Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*. In ed. E. Cannan. New York: Modern Library, 1937.
Sraffa, P. 1951. Editorial introduction to vol. I of *The works and correspondence of David Ricardo*. Cambridge: Cambridge University Press.
Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.

Wages, Real and Money

Henry Phelps Brown

Our knowledge of the movements of real wages from the 12th century onwards in southern England, and later in western Europe, shows them as dominated by the varying pressure of population. A great rise of the English population in the 12th and 13th centuries brought output per head down to a level so low that a bad harvest brought famine. But the Black Death may well have reduced the population by a third, and its periodic recurrence kept numbers down. It was followed by a remarkable rise in the real wage: as indicated by the purchasing power of the wage-rate of the building craftsman or labourer, the real wage rose early in the 15th century to a high level that it sustained for a hundred years. This plateau was as high as any level to be reached, according to the same limited measure, until the second half of the 19th century. But in the 16th century the renewed pressure of population brought food output per head and real wages down again, so that after 1600 the index was running at about half its level of a century before, and in the wage-earner's world famine was recurrent again. A slow but persistent trend of recovery set in after the Civil War. At the mid-18th century a rise of standards of living conspicuous to contemporaries coincided with the onset of a new growth of population, but also an agrarian revolution, which mitigated the fall in real wages that was caused by the Napoleonic Wars.

It was at this time that economists developed a systematic account of the determination of real and monetary wages. Their analysis of supply and demand by way of the subsistence level and the amount of capital or the wages fund, though purporting to be abstract and rigorous, was effectively transcribed from current circumstances. Real wages were seen to be pressed down perpetually towards the level of subsistence by the gravitational force of population pressure. If they were above that level, the propensity of mankind to marry early and beget numerous children would before long increase the number seeking maintenance from 'the funds destined for the employment of labour', that is, the wages fund. If the real wage fell below the subsistence level, marriages would be put off, and hunger and disease would take their toll of existing families, until the wages fund, spread over a diminished labour force, provided a real wage at the subsistence level again. The political economists saw that most payments of wages gave the worker the means of immediate consumption well before the product of his own week's work became available; the wages fund was there in effect to make loans. They went on hastily to assume that the fund could be increased only by the savings of the propertied classes; and inferred that this was the only way in which real wages in the aggregate could be raised. Trade unions could raise the real wages of one group only at the expense of the rest. It was when John Stuart Mill (1869) opened his eyes to the situation of the employer, who could at need divert part of his cash flow to swell his pay packets, that he abandoned the notion of a distinct and predetermined wages fund.

That the political economists were so hasty in taking up that notion may be explained by their preoccupation with population pressure as much the more variable and disturbing term in the equation. But from the mid-19th century onwards it began to appear that advances in productivity more than offset that pressure: despite the continuing rise in numbers, real wages rose. It was in this setting that economists began to derive the real wage from the marginal product. This application of the marginal analysis of demand to a factor of production was made notably by J.B. Clark (1899) and Alfred Marshall (1890). The principle that the real wage tends to equal the marginal physical product has a firm theoretical basis, in that costs will be minimized when the marginal value products of all factors are proportioned to their prices, and profits will be maximized when these two variables are equal. The marginal productivity theory of wages therefore rests upon belief in the pervasiveness and effectiveness of the drives towards costs minimization and profit maximization in the real world. It has been objected that the theory credits the employer with accurate knowledge, and with an ability to vary the intake of factors, that empirical studies show he lacks: actual employers do not approach wages and employment as the textbooks suppose. But our view should not be bounded by the fixity of circumstances in the short run. Adjustments take place by degrees, or by reconstruction, so that in the course of time the structure of production partakes of the flexibility with which the theory of marginal productivity endows it. As these changes come about, those employers who fail, explicitly or implicitly, to apply the marginal calculus are exposed to the risk of supersession by more efficient competitors. Where any one employer combines labour with other factors in fixed proportions, the variation of proportions may be effected by the choice of buyers between different outputs. The elements, therefore, that the theory of marginal productivity sees as determining the real wage can be accepted as shaping the path that real wages tend to follow through time; though the actual adjustments made within any one firm or the market may be less for years together than the theory calls for.

There is this further reason for accepting the marginal productivity theory, that it provides a hypothesis on which we can account for the movement of real wages and associated variables in a number of Western countries since 1860 (Phelps Brown and Browne 1968; Samuelson 1980, fig. 27–3). Real wages rose, or failed to rise, in proportion to the movements of productivity, that is, of output per head in the whole occupied population. The rate of profit varied cyclically within a band that lay about 10 per cent and itself showed no trend. The capital/output ratio lay within a band whose midpoint was about 2.5. It was implicit in the last two observations that the division of the net product between labour as a whole and capital showed no trend, though it varied cyclically; but it was subject to a displacement through the two world wars. These regularities challenge explanation. It has been shown that they would arise if enterprise and investment were at work pervasively within an aggregate production function that is linear and homogeneous of the first degree, such as the Cobb-Douglas function, on two assumptions-that factors are remunerated according to their marginal products, and that the yield of the function with given inputs is progressively increased over time by a technical progress factor (Solow 1957). With such a function, the marginal product of labour will be a constant proportion of its average product, that is, of productivity, as the inputs of labour and capital vary. The division of the product between labour and capital will also be constant. But to account for the stability of the rate of profit and of the capital/output ratio we have to introduce the technical progress factor, and the observation that this has operated, as a fact of history, so as to produce the degree of stability observed.

The mind may boggle at the degree of simplification required by an aggregate production function. There is attraction, therefore, in proceeding more realistically, and simply fixing upon the endeavour of employers generally to secure a normal profit, under the pressures of bargaining with their workers and competition among themselves. This endeavour acts to stabilize the rate of return on capital. If we add, again as a fact of

history, stability of the capital/output ratio thanks to technical progress, we account for stability of the proportionate shares of capital and labour. But unless we introduce some internal equilibrating process, such as the marginal productivity principle, we cannot account for stability in the relation between the real wage and productivity: and of all the stable relations observed, this is the most unmistakeable.

Three factors must be added to that basic relation. One concerns the varying terms on which the industrial worker can obtain foodstuffs and raw materials. There have been long cycles in the terms of trade between factory and farm, or plantation; when agricultural depression has coincided with industrial, cheap food has kept real wages up at a time of low money wages. The second factor is the ability of trade unions to raise real wages by shifting the distribution of the product. The share of profits was generally smaller after both world wars-after the first, through the resistance of money wages to the deflationary pressure of 1920–22; after the second, through the cost push exerted by the trade unions. This latter European trend was contrasted with the lower trend of real wages in the United States, which was linked with the greater expansion of employment there. Thirdly, in recent years the increase of forms of social benefit, accruing to the household of the wage-earner in his or her capacity as a citizen, has given rise to the concept of 'the social wage'. This together with the increased incidence of direct taxation on wage-earners is taken into account in reckoning changes in their standard of living, which may diverge from those in real earnings from employment.

The course of money wages, as that can be traced mainly in the rates of building workers in Southern England from the 13th century onwards (Phelps Brown and Hopkins 1981), has two striking features-the absence of falls, and the presence of long periods of constancy in the prevailing rate. Both may be ascribed to the power of custom; but there was probably also latent if not overt trade unionism, ready to come into action to resist any attempt to cut rates. The constancy of the wage-rate is the more noteworthy when the cost of foodstuffs fluctuated widely with the harvest from one year to another: it was common for the food prices of the dearest year within any decade to be near double those of the cheapest. Since the ratchet effect stopped the wage-rate from falling back, rises cumulated. By the end of the Napoleonic Wars there had been three major lifts. The first came as a result of the Black Death: the labour force was now reduced, probably by a third, over against undiminished resources in land and the same stock of money. After a hundred years of extraordinarily level trend in prices and money wages came the second lift, in the great Tudor inflation. Through the debasement of the currency and the cheapening of silver this doubled the wage rate between 1532 and 1580; but it has been noted above that through the same century food prices rose so much more than the monetary factory alone would have raised them, as to transmit to real wages the severe reduction demanded by population pressure. The third major lift came, again from the monetary side and in the presence of rising population, through the Napoleonic Wars.

With the extension of industry, and the appearance of data for more countries, in the course of the 19th century, a new pattern appears, that of the eight-year trade cycle. The movement of British money wages in this cycle was summarized by Phillips (1958) in a curve which was taken by many economists to show that the general level rose and fell in accordance with the balance of aggregate supply and demand in the labour market, as that was shown by the rate of unemployment. This provided an equation for the endogeneous determination of what it seemed could otherwise be taken only as imposed from without, by historical forces. But a given rate of unemployment had been associated with quite different wage changes, according as it occurred in the rising or falling phases of the cycle; and in recent years regions with different rates of unemployment had obtained much the same wage rise. It seemed likely therefore that unemployment served not as a measure of the balance of supply and demand, but as an indicator of the phase and intensity of the cycle, and the associated expectations. With these changes of phase, on the employers' side, would go changes in the strength

of their resistance to their workers' claims, as the profitability of their operations and their markets' toleration of price rises varied. The view workers took of the reasonableness of changes in money wages during the cycle was influenced by their impression of their employers' current capacity to pay, and by changes in the cost of living.

This cyclical process resulted typically in a rise of five per cent or more in money wages in the rising phase and boom, with the loss of two or three per cent in the ensuing recession, so that, if this were all, money wages would have risen cumulatively over time. But it was not all. Prices in the market for the products of the labour concerned, or for the raw materials it used and the foodstuffs it consumed, were subject to long waves. In phases of downward pressure on prices, the ceiling over the upward thrust of wage-earners was itself being lowered, and the upward trend of money wages might be brought down, or disappear altogether, as it did in Great Britain during the great Victorian depression of the 1870s and 1880s. In phases of rising prices in world markets, money wages too could rise more. Examination of the trends of money wages over longer periods thus displays the ultimate ascendancy of the forces determining the product price level.

At the same time, a normative process has regulated the relation between money wages and prices, so as to keep real wages on a trend parallel to that of productivity. In times of hard market environment, when product prices have had the upper hand, it is money wages that have had to perform the adjustment. There have been periods of some length, in the Victorian depression and the interwar years, when they did not rise at all from end to end, but this was compatible with a substantial rise in real wages. In the years after World War II in which the market environment offered less resistance to the raising of prices, it was possible for cost push to take the initiative, and achieve progressive wage rises, to which prices had to be adjusted. Money wages can rise relatively to product prices according to the rise in productivity and the possibility of reducing profit margins.

The interwar years can be set within the above scheme of forces, but after 1945 the eightyear cycle did not appear again. Instead, an investment boom set in, of unprecedented duration. It was accompanied throughout most of the Western economies by a progressive rise in productivity, and by an experience of sustained demand in product and labour markets that fostered the expectation of its continuance. Employers had little reason to offer costly resistance to claims for rises in money wages when, so far as these were not offset by higher productivity, they could probably be covered as part of a general movement towards higher prices, without loss of business. Analysts used to tracing inflation to excess of monetary demand were eventually driven to recognize that here was an inflation taking its rise in cost push. The remedy was found in incomes policy. In a number of Western countries the endeavour was made to keep the rise of the general level of money wages down to that of productivity by exhortation, by agreement between central organizations of trade unions and of employers, or by restraints and controls imposed by government; these latter might be supported by price controls. But attempted controls all suffered from the lack of an effective sanction agains the strike in a democratic society; and though real wages might be maintained in the end if all wage-earners gave up their claims to higher money wages, the members of any one group could see that if they gave up their own claim they were losing here and now the opportunity of raising their own real wages. The experience of money and real wages rising year by year for more than twenty years imparted an increasing momentum to expectations; about 1969 in a number of European countries the rate of cost push began to rise. But this was just at the time when the great boom began to run out, and unemployment to mount. This brought no immediate or commensurate check to the rise of money wages. Governments wishing to relieve the unemployment of the 1970s were then inhibited from raising effective demand by the fear that the additional purchasing power would simply be absorbed by higher money wages for those already employed. The contrast was drawn between the buoyancy of money wages that resulted at this time from the expectations and cohesion of the European wage-

earners, and the smaller rise in the more dispersed and individualist labour market of the United States. The autonomy of the level of money wages, as a determinant of the price level, presented itself as a major problem of economic helmsmanship.

Bibliography

Clark, J.B. 1899. *The distribution of wealth*. New York/London: Macmillan.

Marshall, A. 1890. *Principles of economics*, 1st ed. London: Macmillan.

Mill, J.S. 1869. Thornton on labour and its claims. In *Essays on economics & society*, Vol. II, ed. J.M. Robson, Toronto/London: University of Toronto Press/Routledge & Kegan Paul, 1967.

Phelps Brown, E.H., and M.H. Browne. 1968. *A century of pay: The course of pay and production in France, Germany, Sweden, the United Kingdom, and the United States of America, 1860–1960*. London: Macmillan.

Phelps Brown, E.H., and S.V. Hopkins. 1981. *A perspective of wages and prices*. London/New York: Methuen.

Phillips, A.W. 1958. The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica* 25(100): 283–299.

Samuelson, P.A. 1980. *Economics*, 11th ed. Tokyo/London: McGraw-Hill, Koga Kusha.

Solow, R.M. 1957. Technical change and the aggregate production function. *Review of Economics and Statistics* 39: 312–320.

Wagner, Adolph Heinrich Gotthelf (1835–1917)

Hermann Reich

Adolph Wagner was born in Erlangen (Franconia), the son of a professor of physiology. Studied in Heidelberg and Göttingen and taught between 1858 and 1870 in Vienna, Hamburg, Dorpat (today Tartu, Estonia) and Freiburg (Breisgau). In 1870 he was appointed professor of political economy at the University of Berlin, a position which he was to hold for 46 years, and where he became one of the most important economists of the German Reich.

Wagner tried to steer a middle course between the historical school and its theoretically oriented opponents. At a time when economic theory was neglected in Germany, it was to Wagner's merit that he helped avoid its almost complete disappearance from economic discussion. His important *Grundlegung* (1876) included numerous sociological elements. It emphasized institutional patterns, historical-legal categories and various aspects of the different forms of private and public property. Moreover, he focused on the role of the state, on the psychological motivations of the individual, and on population development, where he had adopted Malthusian views.

Wagner had started his career as an expert on money and banking (Wagner 1857). But his lasting and outstanding achievement was his work on Public Finance (1871–2), which he freed from its previous fiscal administrative orientation and incorporated into the framework of political economy. In this work the role that Wagner assigned to the state was of primary importance-a concept which has to be seen within the wider context of a particular social philosophy which he himself called 'state socialism' (Wagner 1887, 1912). This notion ought to be understood as a specifically German type of social conservatism, based on an organic concept of the state and on the rejection of laissez-faire liberalism. 'State socialism' aimed at the integration of the working classes into the monarchic state and was thus directed against the growing social democratic party. This goal was to be achieved by a gradual transformation of liberal capitalism into a state interventionist economy-on which Wagner took a firmer stand than most of the social-reform oriented German professors, the so-called 'Socialists of the chair' (Kathedersozialisten). He advocated the nationalization of sectors that showed a high degree of monopolization, especially transport, utilities, banking and insurance (Wagner 1887, pp. 43–4). He also favoured the abolition of private real estate property. Moreover, the State was to intervene in the market through a paternalist social policy and a redistributional tax policy. As a result of what he called the social-welfare principle of taxation Wagner advocated progressive

income taxation, wealth, inheritance, luxury and capital gains taxes (Wagner [1871–2] 1880, pp. 282–92). He formulated a historical 'law' of 'growing public and state activities' as a general consequence of cultural development (Wagner [1876], 1892–3, pp. 892–908).

It was characteristic of this ideology that Wagner, who referred to Kaiser Wilhelm I and Bismarck as 'state socialists' (Wagner 1912, p. 24), was at the time a deeply conservative Prussian nationalist. Like many other German conservatives, he was deeply sceptical towards industrial and capitalist development (Wagner 1901) and antisemitic, which attenuated during his later years. Wagner joined Adolf Stöcker's reactionary and antisemitic Christian Social party, became its vice-president and member in the lower house of the Prussian Diet (1882–5), and was an active member of the Evangelical Social Congress. When the Nazis came to power, Wagner was praised as a precursor of National Socialism (Vleugels 1935). However, there is still a far way to go from Wagner's type of conservatism to Nazi fascism.

Selected Works

1857. *Beiträge zur Lehre von den Banken*. Reprinted, Vaduz: Topos, 1977.

1871. Speech on the social question. In *Social reformers: Adam Smith to John Dewey*, ed. D.-O. Wagner. New York: Macmillan, 1934.

1871–2. *Finanzwissenschaft,* 4 vols. Leipzig/Heidelberg: Winter. First published as a revision of Rau, K.H., *Lehrbuch der Finanzwirtschaft*. Vol. 1: *Einleitung. Ordnung der Finanzwirtschaft. Finanzbedarf. Privaterwerb,* 3rd ed, 1883. vol 2: *Gebühren und Allgemeine Steuerlehre*, 1880. For a translation of excerpts, see *Classics in the theory of public finance*, ed. R.A. Musgrave and A.T. Peacock. London/New York: Macmillan, 1958, 1–15. vol 3 and 4: *Specielle Steuerlehre*, 1889 and 1899–1901.

1876. *Grundlegung der politischen Oekonomie*. Pt. I: *Grundlagen der Volkswirthschaft*; 3rd ed. Leipzig: Winter, 1892–3.

1886. Systematische Nationalökonomie. *Jahrbücher für Nationalökonomie und Statistik* NF 12. Partially translated as 'Wagner on the present state of political economy'. *Quarterly Journal of Economics* 1886.

1887. *Finanzwissenschaft und Staatssozialismus*. Sozialökonomische Texte No. 15, ed. A. Skalweit. Frankfurt: Vittorio Klostermann.

1891. Marshall's Principles of Economics (Book Review). *Quarterly Journal of Economics*.

1901. *Agrar- und Industriestaat,* 2nd ed. Jena: Gustav Fischer, 1902.

1907–9. *Theoretische Sozialökonomik,* 2 vols. Leipzig: Winter.

1912. *Die Strömungen in der Sozialpolitik und der Katheder- und Staatssozialismus*. Berlin: Volkstümliche Bücherei.

References

Clark, E.A. 1940. Adolph Wagner: From national economist to national socialist. *Political Science Quarterly* 55: 378–411.

Heilman, M. 1980. *Adolph Wagner – ein deutscher Nationalökonom im Urteil der Zeit*. Frankfurt/New York: Campus.

Hutter, M. 1982. Early contributions to law and economics: Adolph Wagner's 'Grundlegung'. *Journal of Economic Issues* 16(1): 131–147.

Vleugels, W. 1935. Adolph Wagner: Gedenkworte zur hundertsten Wiederkehr des Geburtstages eines deutschen Sozialisten. *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich*.

Waiting

K. H. Hennings

The term 'waiting' was introduced by MacVane (1887) to replace the term 'abstinence' used by earlier economists. Both terms are so closely related that they will be discussed together.

Despite some misgivings, Senior adopted the term 'abstinence' because 'there is no familiar term to express the act, the conduct, of which

profit is the reward, and which bears the same relation to profit which labour does to wages' (1836, p. 89). The idea that saving implies to abstain from the use of existing goods for consumption purposes had earlier been expressed by Adam Smith ([1759] 1976, pp. 189–90), T.R. Malthus (1820, p. 314), G.P. Scrope (1833, p. 146) and especially John Rae, who had argued explicitly that men 'sacrifice a certain amount of present goods to obtain another, greater amount of goods at some future period' (1934, p. 119). Senior, however, combined saving with investing, and denoted by 'abstinence' both a form of economic activity and its result: 'By the word abstinence we wish to express the agent distinct from labour and the agency of nature the concurrence of which is necessary to the existence of capital' (1836, p. 59); or again 'abstinence expresses both the act of abstaining from the unproductive use of capital, and also the similar conduct of a man who devotes his labour to the production of remote rather than of immediate results' (1836, p. 89). Considered as an 'instrument of production', abstinence was not, however, independent: 'although human labour and the agency of nature independently of that of man are the primary productive powers, they require the concurrence of a third productive principle to give them complete efficiency' (1836, p. 58) because without it time-consuming production is not possible. Abstinence is thus associated by Senior with the idea that production takes time. As 'to abstain from the enjoyment which is in our power, or to seek distant rather than immediate results, are among the most painful exertions of the human will' (1836, p. 60), abstinence commands a price as a scarce factor of production which puts it on a par with the other, primary factors of production.

As a term denoting saving and investing, i.e. a form of economic activity which commands a reward, 'abstinence' was adopted by J.S. Mill, Cairnes and Jevons, Bastiat and Cherbuliez, Hermann and Roscher, and soom became part of established theory. Its general and rapid adoption indicates both the inadequacy and the end of a pure cost of production theory of value. Lassalle castigated and ridiculed it by comparing the sacrifices of millionaires to those of small savers (1864, pp. 110). This critique was answered by Loria (1880, pp. 610–24) and later by Macfarlane (1899) with the argument that the savings of millionaires were intramarginal, and that their rewards benefited from savers' surplus. It was probably for that reason that the term continued to be used for the act of saving and investing. Thus J.B. Clark (1899, p. 134) wrote: 'abstinence is the relinquishment, once and for all, of a certain pleasure of consumption and the aquisition of a wholly new increment of capital'. Yet more and more the term was considered unsatisfactory. Cairnes (1874, pp. 88–95) had unsuccessfully proposed the term 'postponement' in its place. By contrast, MacVane's suggestion (1887) to replace it by 'waiting' was taken up by Marshall, and later by Cassel and others, and subsequently adopted generally.

Marshall equated 'waiting' with 'postponement of enjoyment' (1920, p. 233) or 'saving' (1920, p. 830) and argued that 'the growth of wealth involves in general a deliberate waiting for pleasure which a person has . . . the power of commanding in the immediate present' (1920, p. 234). In a similar way Carver (1893) associated 'abstinence' with the disutility of saving when he argued that the rate of interest is determined jointly by a falling marginal productivity of capital schedule and a rising marginal abstinence schedule. Carver showed also that abstinence is related to, but not the same as, the rate of time preference. Both Marshall and Carver thus distinguished between saving and investing, reserving the term 'waiting' (or 'abstinence') for saving.

Marshall also took up Senior's association of abstinence with time-consuming production, and extended it to consumption. Production, if it takes time, requires waiting because most outputs will appear only after most inputs have gone into the process. Similarly, durable consumer goods involve waiting because their services extend over time. The exertions, efforts and sacrifices involved in such economic behaviour in production Marshall counted among the real costs of production; where durable consumer goods are involved, they were counted among the real benefits of their use (1920, p. 339).

In this form, the 'abstinence theory' was severely criticized by Böhm-Bawerk (1921, vol. I, ch. 9 and appendix pt. 4). Based in effect on Senior's denial that capital was an independent factor of production, Böhm-Bawerk maintained that abstinence or waiting could not be counted among the real costs of production. Instead of adding the rewards for abstinence or waiting to the expenses of production the correct way was in his view to take account of the under-valuation of future benefits as reflected in such rewards and count among the (money) costs of production only the rewards of 'primary' factors.

The debate which followed this critique was obfuscated by Cassel because he changed the meaning of the term 'waiting'. Situating his discussion in the context of a price theory, Cassel identified the 'supply of waiting' with savings, thus changing the emphasis from a form of economic behaviour to its results (in money rather than in real terms). At the same time, Cassel resurrected Senior's association of saving with investing, and abstinence with a factor of production, and identified the 'demand for waiting' with the total money value of capital invested (1903, chs 3 and 4). While Cassel's procedure had the advantage of interpreting the price of waiting as the price of keeping a particular stock of capital in use, his adoption of a money value measure of capital (which was not discussed in detail) was confusing if not confused. It did, however, prepare the ground for the debates about saving and investment and their determinants which dominated macroeconomic discussions in the 1930s. In his later treatise (1918) Cassel tended to use the term 'use of capital' *(Kapitaldisposition)* in place of 'waiting'; this indicates that he discussed other issues than the microeconomic ones which had dominated the debates from Senior to Marshall and Böhm-Bawerk.

From Senior to Marshall, abstinence or waiting was associated primarily with the idea that saving involves sacrificing goods available in the present for consumption in order to invest them, and that profits can be regarded as reward for such economic activity. Yet as Rae had shown before the terms were coined, and Carver showed later, the reward for such intertemporal (re-)allocations depends not only on the characteristics of saving behaviour, but also on the productivity, or profitability, of the investment opportunities open to those willing to save. To that extent, therefore, the notion of abstinence for waiting as the activity which is rewarded by profits is misleading. Nor are such terms required to characterize economic behaviour concerned with intertemporal allocations.

From Senior to Böhm-Bawerk, abstinence or waiting was also associated with the notion that the use of durable goods in production as well as in consumption results in time-consuming economic processes, and that profit or interest is in some sense the reward for or price of the capital tied up in such processes. Yet while it is correct that time-consuming economic processes involve 'waiting' or abstinence from immediate consumption, it is not at all clear why such a characteristic of production processes should be given the status of a factor of production. Insofar as Böhm-Bawerk's critique is pertinent (see Fraser 1937, ch. 14). Nor has it proved possible, in view of the many possible temporal structures such processes can assume, to define measures for the waiting or abstinence involved in them which are such that one can speak of profits or interest as reward or price of such a factor of production (see Haavelmo 1960; Hicks 1979). In spite of recent attempts to revive the notion (see Yeager 1976) 'abstinence' or 'waiting' do not seem to be terms which are useful in economic theory beyond denoting, in a rather general manner, a characteristic feature of time-consuming economic processes.

See Also

- ▶ Abstinence
- ▶ Impatience
- ▶ Time Preference

Bibliography

Bastiat, F. 1850. *Harmonies économiques*. Paris: Guillaumin. *Economic harmonies*, London: Murray, 1860.

Cairnes, J.E. 1874. *Some leading principles of political economy newly expounded*. London: Macmillan.
Carver, T.N. 1893. The place of abstinence in the theory of interest. *Quarterly Journal of Economics* 8: 40–61.
Cassel, G. 1903. *The nature and necessity of interest*. London: Macmillan.
Cassel, G. 1918. *Theoretische Sozialökonomie*. Leipzig: Deichert. 4th edn. 1927. *Theory of social economy*. London: Unwin, 1923. New edn, London, Bern, 1932.
Cherbuliez, A.E. 1862. *Précis de la Science Economique*. Paris: Guillaumin.
Clark, J.B. 1899. *The theory of distribution*. New York: Macmillan.
Fraser, L.M. 1937. *Economic thought and language*. London: Black.
Goss, B.A. 1980. Adam Smith on abstinence. *Australian Economic Papers* 19: 16–21.
Haavelmo, T. 1960. *A study in the theory of investment*. Chicago: Chicago University Press.
Hicks, J.R. 1979. Is interest the price of a factor of production? In *Classics and moderns*, ed. J.R. Hicks, Collected Essays Vol. III, Oxford: Blackwell, 1983.
Jevons, W.S. 1871. *The theory of political economy*. London: Macmillan, 1970.
Lassalle, F. 1864. *Herr Bastiat-Schultze von Delitzsch der ükonomische Julian, order Kapital und Arbeit*. Berlin: Schlingmann.
Loria, A. 1880. *La rendita fondiaria*. Milan: Hoepli.
Macfarlane, C.W. 1899. *Value and distribution*. Philadelphia: Lippincott.
MacVane, S.M. 1887. Analysis of cost of production. *Quarterly Journal of Economics* 1: 481–487.
Malthus, T.R. 1820. *Principles of political economy*, 2nd ed. London: Murray, 1836.
Marshall, A. 1920. *Principles of economics*. 8th edn. 9th (Variorum) edn, London: Macmillan, 1961.
Mill, J.S. 1848. *Principles of political economy*. London: Parker.
Rae, J. 1934. *Statement of some new principles on the subject of political economy*. Boston: Hilliard/Gray and Co.
Roscher, W. 1854. *Grundlagen der Nationalökonomie*. Stuttgart: Cotta.
Scrope, G.P. 1833. *Principles of political economy*. London: Longman/Rees/Orme/Brown/Green/Longman.
Senior, W.N. 1836. *Outlines of the science of political economy*. London: Longman, 1850.
Smith, A. 1759. *Theory of moral sentiments*. Oxford: Clarendon, 1976.
von Böhm-Bawerk, E. 1921. *Kapital und Kapitalzins. Erste Abteilung: Geschichte und Kritik der Kapitalzins theorien*. 4th edn, Jena: Fisher. Trans. as Capital and interest, Vol. 1: History and critique of interest theories. South Holland: Libertarian Press, 1959.
von Hermann, F.B.W. 1874. *Staatswirtschaftliche Untersuchungen*, 2nd ed. Münich: Ackermann.
Yeager, L.B. 1976. Toward understanding some paradoxes in capital theory. *Economic Inquiry* 14: 313–346.

Wakefield, Edward Gibbon (1796–1862)

Donald Winch

Keywords
Colonization; Exploitation; Imperialism; International migration; Wakefield, E. G.

JEL Classifications
B31

Wakefield was born in London on 20 March 1796, the eldest son of Edward Wakefield, a radical Quaker philanthropist, statistician, and author of a standard work on Ireland which was highly regarded by Ricardo, James Mill and other members of the philosophic radical circle. His son was to become one of the more colourful characters to inhabit the margins of the history of economic debate, and can be variously described as a publicist, politician and author. Apart from his practical and frequently controversial contributions to the development of Australia, Canada and New Zealand, he left a distinctive mark in the annals of classical political economy during the middle third of the 19th century.

After a chequered education at Westminster School and Edinburgh High School, from which he was expelled in 1811, Wakefield first read for the Bar and later served as secretary to the British envoy to the Court of Turin (1814–20). In 1816 he successfully eloped with a 16-year-old Ward-in-Chancery who died in childbirth in 1820. From 1820 to 1825 he served with the British legation in Paris and entertained ambitions of entering Parliament. In 1826 he made an attempt to acquire a rich wife by the most direct means available: he abducted the daughter of a wealthy family from her school and married her at Gretna Green. He was apprehended by her family at Calais and subsequently given a three-year sentence, which he

spent studying capital punishment and transportation, writing a powerful pamphlet condemning the former, and turning the latter into what was to become a lifetime's preoccupation with colonization. His first work on the subject, *A Letter from Sydney* (1829), purporting to be the reflections of a disillusioned settler on the poor prospects for Australian social and economic development, was actually written from Newgate prison. After his release Wakefield produced a spate of books, articles and prospectuses on the subject of colonization which led to the formation of the National Colonization Society in 1830 – a society which obtained the support of a number of Members of Parliament and of the youthful John Stuart Mill. Although most of his writings dealt with colonization in one form or another, his work on *England and America*: *A Comparison of the Social and Political State of Both Nations* (1833) is of wider interest for its diagnosis of the cause of the 'uneasiness of the middle classes' and for its economic interpretation of slavery. Wakefield also produced an edition of the *Wealth of Nations* (1835–9) which has some interesting editorial comments.

Wakefield's views on colonization were based on a dual analysis of Britain's need for an outlet for its surplus capital and population and a diagnosis of the causes of weak economic development in colonies of new settlement enjoying access to abundant land. His own schemes for 'systematic colonization' were intended as an almost self-regulating solution to both of these problems. Making use of ideas derived from the work of Robert Gourlay, Wakefield advanced a theory of growth in new countries which was designed to support a plan of optimal development. Contrary to the received view, he maintained that access to free or cheap land was responsible for population dispersion, scarcity of labour for hire, and consequent inability to reap the benefits of economies of scale through market concentration and the combined efforts of capital and labour. Under these circumstances the 'natural' pattern of development led to stagnation. Convict labour in Australia and slavery in the American South were both unsatisfactory expedients adopted to deal with a problem that could only be overcome by charging a 'sufficient price' for public or waste land which would deter premature dispersion, stabilize a revolving wage-labour force, and create a fund that could be used to subsidize immigration. The price was defined as one that was high enough to delay land acquisition by newly arrived immigrants without capital of their own, and low enough not to discourage voluntary immigration by reducing real wages and the return on capital.

Colonization on this plan required a new beginning in a colony that was not contaminated by convict labour; and for this purpose Wakefield initially chose South Australia, forming an association for this purpose in 1834. When his proposals were diluted in operation by the founders of the colony (among them another political economist, Robert Torrens), Wakefield turned his attention to New Zealand, serving as the Director of the New Zealand Colonization Company from 1839 to 1846. In 1838 he accompanied Lord Durham on his mission to Canada and wrote the appendix on land disposal to the resulting Durham report.

Wakefield's ideas are of interest for a number of reasons. He belongs to the non-Ricardian underworld by virtue of his attack on Say's Law, the wage-fund doctrine, and the associated idea that capital and labour could never be in surplus together – a mirror image of the problem in colonies where both were scarce. Yet his success in convincing John Stuart Mill and other economists of the correctness of his diagnosis of British and colonial problems gave new significance to the export of capital and labour to colonies and hence to the whole subject of colonization and the development of new countries as a topic within orthodox political economy.

Wakefield also plays a part in the Marxian tradition, or rather its demonology, as a result of Marx's decision to devote a chapter of *Capital* (vol. 1, ch. 23) to showing how Wakefield, under colonial conditions of labour scarcity, had been forced to reveal the underlying logic of capitalist exploitation. What could be achieved quite naturally under European conditions had to be created artificially in new colonies, with the additional subtlety that having served a term of exploitation, the wage-labourer had to pay for his replacement. One could also claim that Wakefield, less

unwittingly, anticipated Hobson and Lenin in providing an economic interpretation of imperialism as a necessary response to stagnation in mature capitalist economies.

In 1853 Wakefield finally practised what he had been preaching by emigrating to New Zealand, where he died in 1862.

Selected Works

1968. *The collected works of Edward Gibbon Wakefield*, ed. M.F. Lloyd Prichard. London: Collins.

Bibliography

Mills, R.C. 1915. *The colonization of Australia, 1829–42: The Wakefield experiment in expire building*. London: Sidgwick & Jackson.

Winch, D. 1965. *Classical political economy and colonies*. London: Bell & Sons.

Wald, Abraham (1902–1950)

E. R. Weintraub

Keywords

Cassel, G.; General equilibrium; Menger, K.; Schlesinger, K.; Sequential analysis; Statistical decision theory; Wald, A.

JEL Classifications

B31

Born in Cluj, Rumania, Wald came to Vienna in 1927 to study mathematics with Karl Menger, the geometer and son of the economist Carl Menger. Menger introduced Wald to the active mathematical group in Vienna, and secured for him a position as mathematical tutor to the economist Karl Schlesinger. This led to Wald's producing the first proofs of existence for models of general equilibrium; his analysis was based on Cassel's restatement of the Walrasian model, as modified by Schlesinger's treatment of free goods. These works were published in the proceedings of Menger's mathematical colloquium, and a summary was published in the *Zeitschrift für Nationalökonomie* in 1936. These papers were remarkable for their time and, with von Neumann's paper on equilibrium in a model of an expanding economy, are the first significant contributions to the mathematical analysis of general equilibrium models in economics. Wald is the link between the early work by Walras and the later work by Kenneth Arrow, Gerard Debreu and Lionel McKenzie on the existence of competitive equilibria.

A fine mathematician, Wald was nevertheless prevented from gaining a regular academic position because of Viennese anti-Semitism. Menger helped Wald secure a consultancy position with Oskar Morgenstern who directed the Institut für Konjunkturforschung, where Wald took an interest in the statistical problems that were associated with the analysis of business cycles. Wald's book on seasonal adjustment of time series was a result of his work at Morgenstern's Institut.

Wald was able to escape from Vienna when the Nazis arrived, and made his way to the United States where he initially secured a fellowship, in 1938, at the Cowles Commission which was then at Colorado Springs. When the Commission moved to Chicago, Wald obtained a position, on a Carnegie grant, as Harold Hotelling's assistant at Columbia University. He moved to a faculty post at Columbia in 1941, and was promoted to Associate Professor in 1943 and Professor in 1944.

Wald's contributions to statistics are immense. His most significant paper appeared in 1939 in the *Annals of Mathematical Statistics* as 'Contributions to the Theory of Statistical Estimation and Testing Hypotheses' (in Wald, 1955). This paper, written before modern decision theory was developed, contained notions of decision space, weight and risk functions, and minimax solution (based on von Neumann's 1928 paper on game theory). Wald's paper was not appreciated at the time, much as was the case with his papers on general equilibrium theory. He did not return to statistical

decision theory until 1946, after von Neumann and Morgenstern had presented the theory of games.

During the Second World War, Wald worked with the Statistical Research Group and developed much of the theory of sequential analysis. Although he did not create the idea of taking observations sequentially, Wald did invent the sequential probability ratio test. This original material was published in 1947 after wartime restrictions were lifted.

In 1950, at the height of his powers, Wald and his wife died in a plane crash in India.

See Also

- ▶ Existence of General Equilibrium
- ▶ Revealed Preference Theory
- ▶ Statistical Decision Theory

Selected Works

In 1952 *The Annals of Mathematical Statistics* devoted the first part of its volume 23 to a memorial to Wald. Articles on Wald by Jacob Wolfowitz, Karl Menger and Gerhard Tintner were followed by a complete bibliography of Wald's writings. Wald's professional correspondence, and papers from his Viennese days, cannot be located, though it is possible that Karl Menger's archives, currently closed to examination, may contain some material on Wald.

1934. Über die eindeutige positive Lösbarkeit der neuen Produktions gleichungen I. In *Ergebnisse eines mathematischen Kolloquiums*, 1933–-34, ed. K. Menger. Trans. W. Baumol as 'On the unique non-negative solvability of the new production equations, Part I', in *Precursors in Mathematical Economics*, ed. W.J. Baumol and S.M. Goldfeld, London School of Economics Series of Reprints of Scarce Works on Political Economy No. 19, London: London School of Economics, 1968.

1935. Über die Produktionsgleichungen der ökonomischen Wertlehre II. In *Ergebnisse eines mathematischen Kolloquiums*, ed. K. Menger, 1934–35. Trans. W. Baumol as 'On the production equations of economic value theory, Part II', in *Precursors in Mathematical Economics*, ed. W.-J. Baumol and S.M. Goldfeld, London School of Economics Series of Reprints of Scarce Works on Political Economy No. 19, London: London School of Economics, 1968.

1936. Über einige Gleichungssysteme der mathematischen Ökonomie. *Zeitschrift für Nationalökonomie*. Trans. O. Eckstein as 'On some systems of equations in mathematical economics', *Econometrica* 19 (1951), 368–403.

1947. *Sequential analysis*. New York: Wiley.

1950. *Statistical decision functions*. New York: Wiley.

1955. *Selected papers in statistics and probability*. New York: McGraw-Hill.

Walker, Francis Amasa (1840–1897)

A. W. Coats

Keywords

Bimetallism; Residual claimant theory of wages; Wages fund; Walker, F. A.

JEL Classifications

B31

Internationally the most widely known and esteemed American economist of his generation, Walker had a varied and distinguished public career. After obtaining his AB at Amherst in 1860, he studied law for one year before joining the Northern army and was successively a Civil War general, deputy to David A. Wells in the Budget Office, chief of the US Treasury's Bureau of Statistics, Superintendent of the Census of 1870 and 1880, Professor of Political Economy and History at Yale's Sheffield Scientific School, and also occasionally at Johns

Hopkins, and President of the Massachusetts Institute of Technology, 1881–97. At home Walker was primarily known as an outstanding educational administrator and statistician, for he permanently raised the standards of government statistics, helped to create a permanent Bureau of the Census, and served as President of the American Statistical Association from 1882–97. Abroad, he was recognized more as an economic theorist, especially for his work on wages, money and currency policy.

His attack on the wages fund and formulation of a residual claimant theory of wages attracted widespread attention, though it gained few adherents. His writings on money, and a textbook on political economy, were also well regarded, and his support for bimetallism, which involved the monetization of silver, represented an important contribution to a highly controversial current policy debate. In 1878 Walker was appointed US Commissioner to the Paris International Monetary Conference, but in later years he refused comparable invitations as he became disenchanted with the slow progress of international negotiations.

A moderate critic of the ruling classical laissez-faire orthodoxy, Walker responded sympathetically to the rising young generation of German-trained American economists, hence he was both an obvious and in practice ideal choice as first President of the American Economic Association, from 1885 to 1892. His presidential addresses provide revealing insights into the condition of the subject and the emerging economics profession during those critical years. Walker was an open-minded man, forthright in expression but fair in controversy. He believed in competition while recognizing its imperfections. An undoctrinaire free trader, he was concerned about the growth of immigration and the decline in the native birth rate. An advocate of moderate reductions in hours of work, he was one of the first American economists to recognize entrepreneurial gains as rents of ability. Fifty years after his death, Walker's eminence was acknowledged when the American Economic Association selected his name for its most distinguished award, the Walker Medal.

Selected Works

1876. *The wages question*: *A treatise on wages and the wages class*. New York: H. Holt, 1906.
1878. *Money*. London: Macmillan; New York: H. Holt, 1891.
1879. *Money in its relations to trade and industry*. New York: H. Holt, 1907.
1883a. *Land and its rent*. Boston: Little, Brown, 1891.
1883b. *Political economy*, 3rd revised and enlarged ed., 1911. Abridged, 1912. New York: H. Holt.
1896. *International bimetallism*. London: Macmillan; New York: H. Holt, 1897.
1899a. *Discussions in economics and statistics*, 2 vols, ed. D.R. Dewey. New York: H. Holt.
1899b. *Discussions in education*, ed. J.P. Munroe. New York: H. Holt.

Bibliography

Newton, B. 1968. *The economics of Francis Amasa Walker: American economics in transition*. New York: Kelley.

Wallace, Alfred Russel (1823–1913)

Mason Gaffney

Wallace discovered independently the principle of natural selection which he and Charles Darwin co-published in 1858. Wallace's speciality was zoogeography, consonant with his early work as a surveyor and his later on land policy. 'Wallace's Line' through the Makassar Straits dividing Indo-Malayan from Austro-Malayan fauna is an enduring, if modified monument. His *Geographical Distribution of Animals* (1876) has dominated the field.

The acclaim given to natural selection gave new authority to the natural scientist, shaping ideas of social conduct based on new insights into man and nature. For Wallace it was *land*, man and nature. He was influenced by the young

Herbert Spencer of *Social Statics* (1850) who wrote then of equal rights to land.

'Survival of the fittest' (Spencer's apothegm) lent itself easily to legitimize predation among humans, reinforcing the dismal doctrines of Malthus. 'Social Darwinism' as articulated by the later Spencer, T. H. Huxley, W. G. Summer and others became a materialistic, evolutionminded elitism. But Wallace saw mental, social and spiritual factors guiding human evolution. He put his scientist's prestige on the popular side of social issues.

Land policy was aflame with strife. Wallace was outraged by the clearances of the times, and past enclosures, and Irish landlordism, and slums where evictees huddled. In *The Malay Archipelago* (1869) he digressed from natural science to laud primitives as civilized, and score Britain as barbaric. John Stuart Mill sought Wallace out to join the Land Tenure Reform Association which occupied Mill's last years, 1871–3.

Mill's object was to nationalize future increments of land value (or perhaps of rent). Wallace deferred to Mill, but later grew more radical, moved by the Irish land agitation. In 1880 he criticized Parnell's programme for Irish peasant proprietorship as not abolishing privilege, but merely reshuffling some titles.

In 1881 The Land Nationalization Society was formed on Wallace's lines, with him as president. In *Land Nationalization* (1882) he laid out his programme. The state was to assume title to all land, compensating holders with an annuity for the duration of lives in being, based on the same net income from the land derived before nationalization. All men could now lease parcels for use, consummating the natural relation of man to nature, alternating between industry and agriculture.

Land nationalization was not collectivist. Lessees were to have secure tenure and tenantrights to improvements. Rents to the state would be based on the assessed 'inherent value', dependent only on natural conditions. As a surveyor and a biogeographer Wallace readily distinguished inherent value from man's improvements to land, which he saw as transitory.

Present holders would lose the right to sell; to bequeath; and to let land. They could only hold what they occupied and used themselves. Wallace saw land inheritance as a dysgenic factor in human evolution, giving an artificial advantage to unfit heirs both individually and in their collective power to control social evolution.

Wallace held that man's mind overrode the action of natural selection on his body. The mind understood and controlled natural forces. Without inheritance natural selection would be based on individual merit. Universal education would delay marriage; social reform would reduce male death rates, and female choice would replace Malthusian frightfulness as the engine of selection to improve the race.

Wallace's view was kindred in spirit to Henry George's *Progress and Poverty* (1879), although Wallace had less regard for the market. Both saw man as needing land. Their mutual disapproval of Parnellism brought them together, and both submerged methodological differences to further their common concept. Wallace gave him a platform when George toured Britain. Wallace cast George as a theorist who confirmed Wallace's inductive argument, perhaps underrating George's journalistic background. For many years single tax and land nationalization were closely linked by friend and foe. To Liberal Prime Minister Herbert Asquith they were two arms of a pincers, driven together by valuation: 'Tax or Buy' was his slogan.

In later years Wallace went socialist, but continued to support single tax, which from 1895 to 1914 dominated land reform efforts in Britain. But land reform when it came in the Town and Country Planning Act (1947), although neither would have owned it, evinced more Wallace than George.

See Also

► George, Henry (1839–1897)

Selected Works

1882. *Land nationalization*. London: Swan Sonnenschein.

1869. *The malay archipelago*. 2 vols. London: Macmillan.
1905. *My life: A record of events and opinions*. 2 vols. London: Chapman & Hall.

Bibliography

Bannister, R.C. 1979. *Social Darwinism*. Philadelphia: Temple University Press.
Clements, H. 1983. *Alfred Russel Wallace, Biologist and Social Reformer*. London: Hutchinson.
Durant, J.R. 1979. Scientific naturalism and social reform in the thought of Alfred Russel Wallace. *British Journal for the History of Science* 12: 31–58.
Fichman, M. 1981. *Alfred Russel Wallace*. Boston: Twayne Publishers.
Lawrence, E.P. 1957. *Henry George in the British Isles*. East Lansing: Michigan State University Press.

Wallich, Henry Christopher (Born 1914)

Stephen H. Axilrod

Wallich was born in Germany in 1914 and became a citizen of the United States in 1944. He did his graduate work in economics at Harvard University and wrote his doctoral dissertation, later his first published book, on the monetary problems of a small open economy.

The blending of theoretical insights with practical knowledge and policy judgements, begun in his dissertation, has been a hallmark of Wallich's career and work. His early experience and background contributed to this approach. As a young man, he worked first in the export business in Latin America and then as a securities analyst in New York. His tenure as head of the Foreign Research Division of the Federal Reserve Bank of New York in the late 1940s included missions advising less developed countries on central banking and fiscal policy issues.

Wallich's subsequent academic and public careers focused mainly on major domestic and international monetary questions and other practical economic issues, including aspects of tax policy. As Professor of Economics at Yale University from 1951 to 1974, he taught graduate courses in money and banking, published widely, took time off to act first as Assistant to the Secretary of the Treasury and then as a member of the President's Council of Economic Advisers, and wrote (alternately with Milton Friedman and Paul Samuelson) a well-respected column on economics for a national weekly. He became a member of the Board of Governors of the Federal Reserve System in 1974, where his particular responsibilities have largely been in the international area.

Wallich's approach to macroeconomics has encompassed an understanding of both the importance of budgetary and tax policy and the critical role played by money, particularly in the process of containing inflation. He has stressed the desirability of international coordination and cooperation. He has also emphasized the need for an efficient micro economy, with appropriate market incentives to minimize rigidities that might lead to unfavourable macroeconomic tradeoffs as between, say, unemployment and prices. When market incentives were working imperfectly, Wallich has favoured remedial action consistent with retaining the flexibility and efficiencies of a market price system, such as tax incentives to restrain wage increases.

Selected Works

1946. Debt management as an instrument of economic policy. *American Economic Review* 36: 292–310.
1950. *Monetary problems of an export economy*. Cambridge, MA: Harvard University Press.
1955. *Mainsprings of the German revival*. New Haven: Yale University Press. Also in German.
1956. Conservative economic policy. *Yale Review* 46(1): 63–73.
1960. *The cost of freedom*. New York: Harper & Brothers. Also in German, Spanish and Portuguese.
1968. Monetary and fiscal policies in open economies under fixed exchange rates. *Journal of Political Economy* 76(4), Pt II: 951–952.

1971. (With S. Weintraub.) A tax-based income policy. *Journal of Economic Issues* 5(2): 1–19.
1981. *Monetary policy and practice*. Lexington: Lexington Books.

Wal-Mart, Economics Of

Emek Basker

Abstract

Wal-Mart is the largest retailer in the world, with stores in 16 countries (including the United States) and annual revenues exceeding $400 billion. Wal-Mart owes its success primarily to its early and persistent investments in technology. Technology has allowed Wal-Mart not only to grow – adding stores in new markets and adding a broad range of products over the past half century – but also to cut its costs, making it a formidable competitor in almost every retail sector. Wal-Mart's competitive effect lowers prices in local markets, in the process driving some of its competitors to contract or shut down.

Keywords

Big box; Chain store; General merchandise; One-stop shopping; Price competition; Retail; Wal-Mart

JEL Classifications

J21; J63; 033; L11; L25; L81

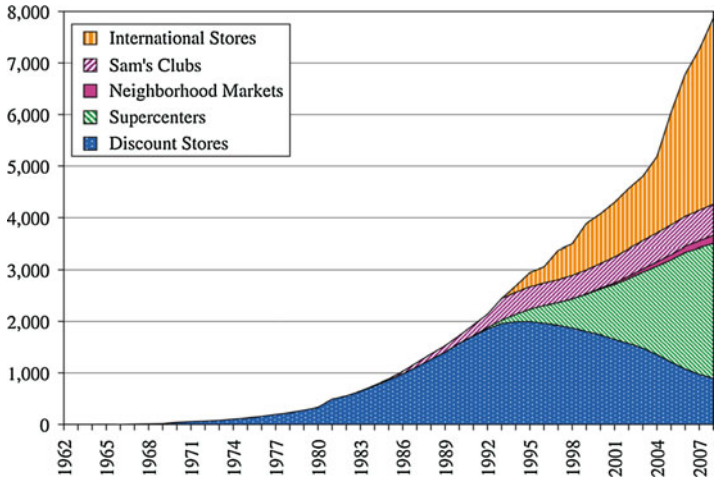
Wal-Mart is a discount general-merchandise retail chain based in the United States, and is the largest retailer in the world. As of 2006, Wal-Mart accounted for 28 per cent of Playtex's sales, 25 per cent of Clorox's, 21 per cent of Revlon's, 13 per cent of Kimberly Clark's, and 17 per cent of Kellogg's (Weinswig and Tang 2006).

Started in 1962 by Sam and Bud Walton as a five-and-dime store in Rogers, Arkansas, the chain had 18 stores when it incorporated in 1970, more than 1700 stores by the end of 1990, and 7873 stores as of 31 January 2009: 891 US Wal-Mart 'Discount Stores' (Wal-Mart's traditional format, selling apparel, housewares, toys, electronics, prescription drugs, and more), 2612 US Supercenters (which include a full grocery store in addition to general merchandise), 602 Sam's Clubs (membership clubs selling a wide range of products), 153 Neighborhood Markets (smaller formats which sell mostly groceries) and 3615 international stores in 15 countries: Argentina, Brazil, Canada, Chile, China, Costa Rica, El Salvador, Guatemala, Honduras, India, Japan, Mexico, Nicaragua, Puerto Rico and the United Kingdom. From 1 February 2008 to 31 January 2009, Wal-Mart's sales exceeded $400 billion; the US divisions alone had revenues exceeding $255 billion, or nearly 1.8 per cent of US GDP. Figure 1 shows the number of stores, by division, at the end of each calendar year (1962–2008).

Wal-Mart's main competitors are Target and Kmart, the two largest US discount general merchandisers, but because it sells such a wide range of products it effectively competes with supermarkets, toy stores, electronics stores, apparel stores and much more. Over 99 per cent of Wal-Mart stores have pharmacies; and most Wal-Mart stores, including 'discount stores', carry at least some groceries.

Not coincidentally, the first Kmart and Target also date to the early 1960s; Wal-Mart's rapid expansion has been emblematic of the widespread rise of chain retailing in the 20th century. Chains accounted for less than 30 per cent of all retail sales in 1948 and over 60 per cent of sales by 2002 (Basker et al. 2008). The growth of retail chains owes much to technology, which has made it possible for a single firm to manage complex supplier relationships, personnel, logistics and distribution.

Popular opinion about Wal-Mart is mixed. The criticisms of Wal-Mart tend to vacillate between two contradictory views: that it competes too aggressively to maximize its profit, for instance by placing extreme conditions on suppliers, aggressively fighting unions, or charging lower prices in markets with heavier competition (see, for example, Norman 2004); and that it uses its



Wal-Mart, Economics Of, Fig. 1 Wal-Mart stores, 1962–2008

market power in non-profit-maximizing ways, such as to reduce access to birth control or to censor music (Bianco 2006, pp. 248–50).

Despite widespread claims of negative employment effects and low wages, research has found very small, and generally positive, effects of Wal-Mart on employment. On the other hand, Wal-Mart has either settled or lost a large number of legal challenges by current and former employees alleging they were required to work off the clock, denied breaks, or denied overtime pay.

Wal-Mart's Advantages: Technology and Scale

Wal-Mart invested in computers early and has continued to make large investments in technology, which accounts for much of its success. In 1969, Wal-Mart installed a computer in its first distribution centre, and it later connected all its stores and distribution centres, along with company headquarters, to a computer network. In the 1980s, Wal-Mart was at the forefront of bar-code adoption (Vance and Scott 1994), just as it is currently a leader in radio frequency identification (RFID) technology, which works by embedding radio transmitters in individual products or cases of products, and reduces much of the cost involved in tracking shipments, inventory and sales. Its Retail Link software, introduced in 1990, connects its stores, distribution centres and suppliers, providing detailed up-to-the-minute inventory data to its suppliers.

These investments have increased Wal-Mart's productivity and made it a formidable competitor. Basker (2007) offers a back-of-the-envelope calculation of Wal-Mart's labour productivity growth compared with the productivity growth in the rest of the retail sector from 1982 to 2002. Wal-Mart's sales per employee grew, in real terms, by 55 per cent over this period; other general merchandisers increased their sales per worker by only 18.5 per cent, and productivity in the rest of the retail sector grew by only 9 per cent.

Economies of scale also play an important role in Wal-Mart's success and its ability to charge low prices to consumers. There is anecdotal evidence that Wal-Mart asks for, and receives, price discounts from suppliers (see Fishman 2006). In addition, the benefit that Wal-Mart is able to squeeze from its investments in technology owes much to its size. Tracking purchases alone would not have enabled Wal-Mart to discover that Strawberry Pop Tarts are a popular item consumers stock before hurricanes (Leonard 2005). A large volume of purchases – from many stores, over a long period of time, in hurricane-prone areas – was also essential.

Finally, there is some evidence that Wal-Mart imports disproportionately to its size, which also

lowers its costs compared with its smaller competitors. But Wal-Mart was not always a major importer. From 1985 to 1992, Wal-Mart gained popular acclaim with its 'Buy American' campaign. This campaign ended abruptly in late 1992 after *Dateline NBC* aired a segment accusing Wal-Mart of producing privatelabel goods in Bangladesh, smuggling textiles into the United States in excess of quotas, and placing imported clothes on racks marked 'Made in the USA' (Gladstone 1992). By 2004, Wal-Mart's imports from China were valued at $18 billion, or 15.4 per cent of US imports of consumer goods from China, more than twice Wal-Mart's share of retail sales (Basker and Van 2008).

Local Effects

An immediate effect of Wal-Mart's entry into a local market is increased price competition. Estimates of Wal-Mart's food prices range from 10 per cent below the competition (Basker and Noel 2009) to 30 per cent below the competition (Hausman and Leibtag 2007b). Basker and Noel (2009) estimate that competing supermarkets and grocery stores cut their prices by about 1 per cent on average when a Wal-Mart Supercenter opens in town. Hausman and Leibtag (2007a) calculate that prices paid by consumers fall by 3 per cent on average, accounting for product and outlet substitution. An earlier study focusing on prices of drugstore products such as shampoo and toothpaste found that these also fall (Basker 2005b).

Increased price competition reduces profits at some incumbent stores and may cause them to contract or exit. On average, almost as many people lose their jobs at other retail establishments as are hired by a new Wal-Mart store. Using publicly available data on Wal-Mart stores' opening dates to 1999, Basker (2005a) estimated that the number of retail jobs in a county increases by 100 the year Wal-Mart opens a new store (relative to what would have happened had Wal-Mart stayed out of the county), and by 50 after five years. Since the average Wal-Mart store over the period of the study employed about 250 workers, this estimate implies that approximately 200 workers at other stores lose their jobs. In addition, the number of wholesale jobs declines by about 30 in the long run, reflecting the fact that Wal-Mart is vertically integrated: unlike the merchants it replaces, Wal-Mart does not rely on local wholesalers. Drewianka and Johnson (2009) find somewhat larger positive effects on employment in general merchandising. In contrast, Neumark et al. (2008) estimate net job losses of 150 workers per Wal-Mart store (implying that 400 workers lost their jobs, on average, as a result of a new Wal-Mart store hiring 250 workers).

Competitors whose profits fall sufficiently eventually shut down. Basker (2005a) estimates a net closing of about five stores in a county as a result of Wal-Mart's opening. Jia (2008) finds a net reduction of two or three small general-merchandise stores, such as dollar and variety stores, if Wal-Mart or Kmart enter a market, compared with if either of these large retailers stays out. Drewianka and Johnson (2009) find little effect on the number of competitors.

The wide range of estimates of Wal-Mart's effect on prices, employment and competitors reflects a fundamental problem of identification: that is, disentangling cause and effect. Basker (2005a, b) addresses this problem by exploiting the time lag between Wal-Mart's initial decision to open a store and its actual opening. Drewianka and Johnson (2009) instead control for pre-existing trends in the outcome variables of interest (employment, wages and so on). Finally, Neumark et al. (2008) use Wal-Mart's geographic pattern of expansion to predict when Wal-Mart will open in each location. These diverse identification strategies – and the different estimates they produce – are the source of controversy and debate. In the absence of a criticism-proof identification strategy, the precise impact of Wal-Mart on employment and other outcomes remains somewhat uncertain. For a comparison of the store-planning and geographic instruments, see Basker (2006).

Business Cycles

Wal-Mart's low prices are particularly attractive to consumers during economic troughs. Basker

(2016) shows that Wal-Mart sells 'inferior goods' in the technical sense of the term: goods for which demand increases (relative to trend) when incomes fall. This analysis uses quarterly data on average sales per store at Wal-Mart over a ten-year period; the same analysis shows that Target's products are 'normal', meaning demand for them increases when incomes rise. Using a monthly data set and Granger causality tests, Jantzen et al. (2009) also find that Wal-Mart's sales growth falls when the overall economy's growth rate accelerates.

Conclusion

As the largest, and arguably most efficient, chain retailer in the world, Wal-Mart leads the way in both technological innovation and arousing public opposition. Ultimately, however, Wal-Mart is just one of many retail chains making these investments, made possible by the same advances in computing technologies that have transformed other sectors of the economy.

Bibliography

Basker, E. 2005a. Job creation or destruction? Labor market effects of Wal-Mart expansion. *The Review of Economics and Statistics* 87: 174–183.

Basker, E. 2005b. Selling a cheaper mousetrap: Wal-Mart's effect on retail prices. *Journal of Urban Economics* 58: 203–229.

Basker, E. 2006. When good instruments go bad. Unpublished paper, University of Missouri.

Basker, E. 2007. Causes and consequences of Wal-Mart's expansion. *Journal of Economic Perspectives* 21: 177–198.

Basker, E. 2016. Does Wal-Mart sell inferior goods? *Economic Inquiry*. forthcoming.

Basker, E., and M. Noel. 2009. The evolving food chain: Competitive effects of Wal-Mart's entry into the supermarket industry. *Journal of Economics and Management Strategy* 18: 977–1009.

Basker, E. and Van, P.H. 2008. Putting a smiley face on the dragon: Wal-Mart as catalyst to U.S.-China trade. Unpublished paper, University of Missouri.

Basker, E., S. Klimek, and P.H. Van. 2008. Supersize it: The growth of chains and the rise of the 'big box' retail format. Unpublished paper, University of Missouri.

Bianco, A. 2006. *Wal-Mart: The Bully of Bentonville*. New York: Doubleday.

Drewianka, S. and D. Johnson. 2009. Wal-Mart and local labor markets, 1990–2005. Unpublished paper, University of Wisconsin–Milwaukee.

Fishman, C. 2006. *The Wal-Mart effect*. New York: Penguin.

Gladstone, R. 1992. No. 1 retailer slammed in 'Dateline NBC' show, stock falls. Associated Press, 21 December.

Hausman, J., and E. Leibtag. 2007a. Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart. *Journal of Applied Econometrics* 22: 1157–1177.

Hausman, J., and E. Leibtag. 2007b. Wal-Mart effects and CPI construction. In *Price and productivity measurement*, ed. W.E. Diewert, B.M. Balk, D. Fixler, K.J. Fox, and A.O. Nakamura. Victoria: Trafford Press.

Jantzen, R., D. Pescatrice, and A. Braunstein. 2009. Wal-Mart and the U.S. Economy. *Eastern Economic Journal* 35: 297–308.

Jia, P. 2008. What happens when Wal-Mart comes to town: An empirical analysis of the discount industry. *Econometrica* 76: 1263–1316.

Leonard, D. 2005. The only lifeline was the Wal-Mart. *Fortune* 152: 74 , 3 October.

Neumark, D., J. Zhang, and S. Ciccarella. 2008. The effects of Wal-Mart on local labor markets. *Journal of Urban Economics* 63: 405–430.

Norman, A. 2004. *The Case against Wal-Mart*. Atlantic City: Raphel Marketing.

Vance, S.S., and R.V. Scott. 1994. *Wal-Mart: A history of Sam Walton's retail phenomenon*. New York: Twayne.

Weinswig, D., and Tang, C. 2006. Equity research: Wal-Mart Stores, Inc. (WMT). Citigroup Global Markets, 24 April.

Walras, Antoine Auguste (1801–1866)

Donald A. Walker

Keywords

Land nationalization; Numeraire; Property rights; Scarcity; Walras, A. A.; Walras, L

JEL Classifications

B31

Amateur economist whose writings have received some limited attention, chiefly because some of his views and economic concepts influenced his

son, Léon. Auguste Walras was born in Montpellier, France, on 1 February 1801, and died in Pau on 18 April 1866. He studied at the Ecole Normale of Paris (1820–23); was a tutor in Paris (1823–31); a secondary school teacher (1823, 1831–5); a professor of philosophy first at the Royal College in Lille (1839) and then at the Royal College in Caen (1840–47); and a regional school superintendent (1847–62).

Believing that an understanding of property requires a sound theory of value, Auguste Walras developed the unoriginal and unsatisfactory thesis, primarily on the basis of admittedly metaphysical considerations, that economic value depends upon scarcity (*rareté*). This he defined as the relation between the quantity of a commodity and the number of people that have need for it. He concluded that only scarce things are appropriated and constitute property (Walras 1831). He then argued that natural law dictates that the state, like the individual, has the right to own property, and that land in particular should belong exclusively to society as a whole. Developing an explanation of the current ownership of land, he pointed out that it is a consequence of social institutions and historical events. During the feudal era, it was placed by the king under the suzerainty of individuals in return for their military services, and their descendants subsequently ruled it as public officials. Since the need for their feudal functions has disappeared, so also has their right to the use of land, and they have become parasites who benefit from economic growth without contributing to it. The class struggle is therefore between landowners and the rest of society, and social justice requires that it be resolved in favour of society as a whole. Believing in conciliation and rejecting revolutionary action, he argued that the state should acquire all land by purchasing it, and should rent it to private users. During the period before complete nationalization the increments in pure land rent arising from the progress of society should be taxed away, and there should be heavy taxes on transfers of land. Since individuals have the right to own what they make, taxation of produced wealth, as distinct from rent, is unjust. It is therefore an advantage of land nationalization that the state would be supported by the rent it earns and taxation could be eliminated (Walras 1848). He regarded this proposal as being founded upon scientific analysis and described himself as a socialist, but it is clear that his interpretation and solution of 'the social problem' – the problem of the poverty of the working class during the 19th century – was highly coloured by his normative views and was bourgeois in character.

Auguste Walras also studied the function of precious metals in the growth of social wealth, in the measurement of value, and in exchange; argued that the increase of wealth is the object of economic science; made a distinction between capital and income, and between the market for services and the market for products; introduced the entrepreneur as a person who buys factors of production and sells products; and devised the concept of a *numéraire* (Walras 1849). He did not fully develop these ideas nor integrate them into a theory of economic behaviour. His main concern in all his work was to buttress his theory of property and his solution to the social problem.

Selected Works

1831. *De la nature de la richesse et de l'origine de la valeur.* Paris/Evreux: Alexandre Johanneau/Imprimérie d'Ancelle Fils. Ed. G. Leduc, Paris: Alcan, 1938.

1848. *La vérité sociale, par un travailleur.* Extracts of MS in Leroy (1928).

1849. *Théorie de la richesse sociale our résumé des principes fondamentaux de l'économie politique.* Paris: Guillaumin.

Bibliography

Jaffé, W. 1984. The antecedents and early life of Léon Walras. *History of Political Economy* 16: 1–57. (ed. D.A. Walker).

Leroy, L.M. 1928. *Auguste Walras, économiste; sa vie, son oeuvre*. Paris: Librairie Générale de Droit et de Jurisprudence.

Walras, L. 1908. Un initiateur en économie politique, A.A. Walras. *Revue du Mois* 6: 170–183.

Walras, Léon (1834–1910)

Donald A. Walker

Abstract

Léon Walras was the initiator of models of purely competitive general economic equilibration and equilibrium, of mathematical treatments of them, and of many aspects of microeconomic theory. In his period of maturity as a theoretician, he developed a comprehensive model that includes exchange, production of non-durable goods, production of capital goods, and monetary behaviour. That model features irrevocable disequilibrium behaviour and capital accumulation, and its equilibrium is therefore path dependent. His last theoretical effort, which was a failure but nevertheless very influential, was to try to develop a virtual and therefore path-independent model that would justify his static equation system.

Keywords

American Economic Association; Amstein, H.; Applied economics; Barone, E.; Bimetallism; Bortkiewicz, L. von; Capital accumulation; Capital theory; Cash-balances theory of money; Cassel, K. G.; Certainty; Circular flow; Comparative statics; Competition; Condillac, E. B. de; Consumers; Convergence; Cournot, A. A.; Cousin, V.; Demand function; Durable goods; Dynamic equation system; E commodity; Economic growth; Edgeworth, F. Y.; Entrepreneur; Exchange; Excess demand; Existence of equilibrium; Fiat money; Fisher, I.; Galiani, F.; General equilibrium; Human nature; Immaterial goods; Interrelated markets; Investment; Isnard, A. N.; Jevons, W. S.; Justice; Landlords; Lexis, W.; Losses; Marginal productivity theory; Marginal utility; Marginal utility of money; Marshall, A.; Mathematical economics; Mature comprehensive model; Methodology; Mill, J. S.; Monetary equilibrium; Monopoly; Moore, H. L.; Moving equilibrium; Multi-good exchange; Nationalization of land; Natural monopoly, nationalization of; Non-durable goods; Non-virtual model; Normative economics; Pantaleoni, M.; Pareto, V.; Path dependency; Parameters of the models; Period analysis, Walrasian; Poinsot, L.; Positive economics; Preferences; Price stabilization; Production model; Profit; Proudhon, P.-J.; Quesnay, F.; Rareté; Rate of interest; Real balances; Saving; Say, J.-B.; Schultz, H.; Schumpeter, J.; Services; Simultaneous equations; Smith, A.; Social economics; Stability of equilibrium; Static equation system; Stationary equilibrium; Supply function; Tatonnement; Taxation, injustice of; Turgot, A.R.J.; Uncertainty; Uniqueness of equilibrium; Utility; Utility maximization; Veil of money; Virtual model; Wages of management; Walras, A. A.; Walrasian equilibrium; Walrasian pricing; Walras, L.; Walras's law; Wicksell, J. G. K.; Wicksteed, P. H.; Workers; Written pledges

JEL Classifications

B31

Léon Walras was the founder of models of general economic equilibrium.

Biography

Walras was born on 16 December 1834 in Evreux, which is in the Department of Eure in France, and christened Marie Esprit Léon. His father was Antoine Auguste Walras, a secondary school administrator and an amateur economist; his mother was Louise Aline de Sainte Beuve, the daughter of an Evreux notary. After studying at the College of Caen from 1844 to 1850, he entered the Lycée of Douai, where he received the bachelier-ès-lettres in 1851 and the bachelier-ès-sciences in 1853. He entered the School of Mines of Paris in 1854, but, finding the course of preparation of an engineer not to his liking, he gradually abandoned his academic studies in order to cultivate literature, philosophy and social science. Although those efforts resulted in a short story and a novel, *Francis Sauveur* (1858), it rapidly

became apparent to him that his true interests lay with social science. Accordingly, in 1858 he agreed to his father's request to devote himself to economics and promised to continue his father's investigations (1965, vol. 1, pp. 1–2).

During his youth in Paris, Walras became a journalist for the *Journal des Economistes* and *La Presse* from 1859 to 1862, the author of a refutation on philosophical grounds of the normative economic doctrines of P.-J. Proudhon (Walras 1860), an employee of the directors of the Northern Railway in 1862, and managing director of a cooperative association bank in 1865. He gave public lectures on cooperative associations in 1865; was co-editor and publisher with Léon Say of the journal *Le Travail*, a review devoted largely to the cooperative movement, from 1866 to 1868; and, during those years, gave public lectures on social topics (Walras 1868) in which he advocated Victor Cousin's doctrine of compromise between economic classes. After the failure of the association bank in 1868, he found employment with a private bank until 1870 (1965, vol. 1, pp. 3–4). During the 1860s he tried intermittently to obtain an academic appointment in France, but he lacked the necessary educational credentials, and the 11 economics positions in higher education in France were monopolized by orthodox economists who, he complained, passed their chairs on to their relatives (1965, vol. 1, p. 3). His fortunes ultimately changed as a result of his participation in 1860 in an international congress on taxation in Lausanne, for that drew him to the attention of Louis Ruchonnet, a Swiss politician who secured his appointment in 1870 to an untenured professorship of economics at the Academy (subsequently University) of Lausanne in Switzerland. He was made a tenured professor there in 1871, and held that position throughout his teaching career.

Walras's personal life was initially unconventional. He and Célestine Aline Ferbach (1834–79) formed a common law union in the late 1850s. She had a son, Georges, by a previous liaison, and she and Walras had twin daughters in 1863, one of whom died in infancy. In 1869 he married Célestine, thereby legitimizing their daughter, Marie Aline, and adopted Célestine's son. A long illness of Celestine's and the meagreness of Walras's salary made life very difficult for him for several years. His time and energy were sorely taxed not only by the need to care for his wife but by the need to supplement his salary by teaching extra classes, contributing to the *Gazette de Lausanne* and the *Bibliothèque Universelle*, and working as a consultant for La Suisse insurance company. Five years after Célestine's death in 1879, Walras married Léonide Désirée Mailly (1826–1900). The marriage was a happy one. Her annuity relieved his financial distress, and his situation was further improved in 1892 by an inheritance of 100,000 francs from his mother, which enabled him to pay debts incurred in publishing and disseminating his works, and to buy an annuity of 800 francs.

Influences Upon His Thought

Walras's professional life was devoted to research and teaching. He frequently asserted that his research was a development of his father's and that was true in some respects. It was under the influence of his father's classification of economic studies that Léon, as early as 1862, planned the division of his life's work into the study of pure theory, economic policies and normative goals (Walras to Jules du Mesnil Marigny, 23 December 1862, L 81; the 'L' stands for 'letter', and, like all correspondence cited in this article, the letter is in Walras 1965), the areas of study that were ultimately set forth respectively in the *Eléments d'économie politique pure* (1874, 1877b, 1889, 1896a, 1900, 1926, 1954), the *Etudes d'économie social* (1896b) and the *Etudes d'économie politique appliquée* (1898). Léon adopted his father's classification of the factors of production into the services of labour, land and capital goods, regarding the source of each service as a type of capital. He adopted his father's definitions of capital as wealth that can be used more than once and of income as wealth that can be used only once, and modified his father's vague term 'extensive utility', clarifying it by defining it as the quantity-axis intercept of a market demand curve. The topic of utility had been treated in French thought by

writers such as F. Galiani (a Neapolitan diplomat at Versailles) and E.B. de Condillac, and it was given further development under the name *rareté* by Auguste Walras, who thus bequeathed to Léon an interest in the concept of utility in relation to the value of goods and an awareness of its dependence upon scarcity, an interest that ultimately led him to define *rareté* as marginal utility. Auguste used the word 'numeraire' to mean an abstract unit of account, and Léon adapted the meaning of the word to his purposes. Auguste's philosophy of social justice and his belief in the desirability of nationalizing land were advocated by Léon throughout his adult life. Léon's major economic theories, however, were derived from his own original inspiration and from sources other than his father. Auguste's greatest contributions to Léon's development as an economist were to encourage him to study economics, to suggest that it should be a mathematical science (A.A. Walras 1831, ch. 18; Walras 1965, vol. 1, p. 493), and to give him access to a library of books on economics.

In that library was A.A. Cournot's *Recherches sur les principes mathématiques de la théorie des richesses* (1838), which Léon Walras credited with having demonstrated that economics could and should be expressed in mathematical form (Walras to Cournot, 20 March 1874, L 253; Walras to H.L. Moore, 2 January 1906, L 1614; Walras 1905a). Cournot's work introduced Walras to the mathematical formulation of exchange between two locations, the theory of monopoly and the associated conditions for profit maximization, the analysis of how prices are repeatedly changed in a search for equilibrium in a purely competitive market, and the demonstration of the effect of large numbers of traders upon the determinacy of price, all topics that Walras developed in his own work (1954, pp. 370–72, 434–40, 443). The first demand curve that Walras beheld was Cournot's and he found it immensely suggestive. He was critical of it, however, because he perceived that Cournot's postulate that the quantity demanded of a good is a function only of its own price is inaccurate if more than two goods are exchanged, and that Cournot did not provide a theoretical rationale for the demand function. Those perceptions, Walras observed, were the starting point for his own inquiries (1965, vol. 1, p. 5; 1905a).

Other ingredients that went into the composition of Walras's theories were provided by Adam Smith, John Stuart Mill, François Quesnay, A.R.J. Turgot and Jean-Baptiste Say. Smith had revealed many of the consequences of unfettered competition and had formulated the concept of normal value. Mill had provided a supplement to and reinforcement of Cournot's and Smith's analyses of competitive pricing (Walras to Ladislaus von Bortkiewicz, 27 February 1891, L 999), and also an extension and grand synthesis of classical doctrines that served Walras as a catalyst for critical studies (Walras 1954, pp. 404–5, 411, 419, 423). Quesnay, in his *Tableau économique,* had expressed the concept of a circular flow of income and of the interdependence of the various parts of the economy. Turgot had clearly delineated the idea of the simultaneous and mutually determined general equilibrium of those parts. Say (1836) had suggested the distinction between the capitalist and the entrepreneur had portrayed the entrepreneur as an intermediary between the market for productive services and the market for outputs, and, in that analysis and in his law of markets, had adumbrated the interdependence between the incomes of the factors of production and the demand for goods. Walras sharpened those ideas and made them a fundamental part of his general equilibrium model.

A.N. Isnard's *Traité des richesses* (1781), a book that Léon owned and that may have been in his father's library, was probably an important source of some of Walras's constructions (Jaffé 1969; Klotz 1994). Like Walras, Isnard was interested in determining equilibrium price ratios, set up a system of simultaneous equations of exchange showing the dependence of the value of each good upon the values of the others, stressed the necessity of having as many independent equations as unknowns, and perceived that the use of a numeraire rendered his system determinate. Anticipating Walras's treatment of production, Isnard assumed given ratios of the inputs in a mathematical model and expressed the costs of production in equation form. Also

like Walras, Isnard studied the allocation of capital among different uses, coming to the conclusion, as did Walras, that in equilibrium the net rate of income of different capital goods is the same.

Finally, Louis Poinsot's *Eléments de statique* (1803) exerted a powerful influence upon Walras. He first read that book when he was 19 years old and kept it at his bedside for decades (Walras to Melle Dick May, 23 May 1901, L 1483). Poinsot painted a picture of the mutual interdependence of a vast number of variables, of how the dynamic forces in physical systems eventuate in an equilibrium in which each object is sustained in its path and relative position. Electrified by the implications of Poinsot's work, Walras conceived a magnificent project. He would emulate Poinsot's vision and analysis in reference to the general equilibrium of the economic universe! That he carried out that plan can be inferred from the striking similarity of the form of his work to Poinsot's, with its careful delineation of functional dependences and parameters, its sets of simultaneous equations and its equilibrium conditions.

Equipped, therefore, with ideas that he could take as building blocks and points of departure, with enough geometry and algebra to put together mathematical statements of economic relationships and conditions -his use of calculus in the *Eléments* came after the first edition – and with the explicit objective of developing a mathematical theory of general equilibrium, Walras began his scholarly activity in Lausanne in 1870. In a period of great creativity that lasted until 1878, he developed most of the foundations of the theory of general equilibrium that appeared in the first edition of the *Eléments*. Walras insisted to his publisher that the first part appear in 1874, before the second part (1877b) was completed, because he learned in May of that year that W.S. Jevons had published a mathematical theory of utility and exchange that was similar to his own (J. d'Aulnis de Bourouill to Walras, 4 May, 1874, L 267), and he was anxious to establish the independence of his discoveries and his priority in regard to most of them. For those same reasons, he published four brilliantly original memoirs containing the heart of his theory of general equilibrium during 1874, 1875 and 1876 (Walras 1877a), paid for the costs of publication of his books, and sent copies of them and of his articles to his many correspondents. From 1878 to 1889, Walras significantly extended and refined his theory of general equilibrium (section "The Mature Comprehensive Model of General Economic Equilibrium").

Walras was an extremely conscientious teacher, but he was an uninspiring lecturer (Walras 1965, vol. 2, p. 560), and the students at Lausanne were interested in careers in law, not in economics, so he failed to develop disciples among them. Moreover, he was with increasing frequency afflicted by bouts of mental exhaustion and irritability that made it difficult for him to lecture and to read and write (see Walker 2006a, pp. 183–7). In 1892 he took a leave of absence to regenerate his strength in order to be able to continue teaching, but soon realized he would find the strain of returning to his tasks insupportable and retired in that year, being at that time 58 years of age.

Subsequently Walras's powers waned rapidly. In 1899 and 1900, he tried unsuccessfully to develop a virtual model of general equilibration and equilibrium (section "Walras's Last Theoretical Work"). After 1900 he completely ceased theoretical construction (Walker 2006a, p. 191), but he wrote a few articles in which he restated earlier ideas. In late 1901 and 1902, he made some inconsequential changes to the *Eléments* which were ultimately put into the text of the fourth edition (1900) to produce the 1926 edition, both of them unfortunately called the 'definitive edition' (1900, p. v; 1926, title page). The latter was chosen for translation by William Jaffé (Walras 1954) and thus became the edition that is known in the anglophone world. Walras died on 5 January 1910 in Clarens, Switzerland.

The Mature Comprehensive Model of General Economic Equilibrium

Walras's Subject Matter

Walras recognized that there were imperfectly competitive market structures and developed a

theory of monopoly to take account of an important class of such phenomena (1954, lesson 41). Realizing, however, that the incorporation of non-competitive elements into his general equilibrium model was beyond his powers (1954, p. 256) and believing that a high degree of competition was 'almost universal' and deserved to be treated as the general case (Walras to Ladislaus von Bortkiewicz, 27 February 1891, L999), he devoted most of his energies to working out a comprehensive model of interrelated 'freely competitive' markets, the aspect of his theoretical work with which this entry is concerned. Competition is most effective, he noted, in organized markets, and he assumed that markets are of that type (1954, pp. 83–4), but he also regarded his analysis as applicable in a general way to less highly organized competitive markets (1954, p. 84).

During his period of maturity as a theoretician, Walras modified and extended the model of competitive general equilibrium that he had presented in the first edition of the *Eléments,* constructing what will be called his mature comprehensive model. He presented it in 1889 in the second and greatest edition (and in the third, (1896a), identical to it in regard to the main body of the text). In the following exposition of that model, all the references to Walras (1954) are to passages that appeared in the 1889 edition.

The model is comprehensive in the sense that it deals with exchange, production, capital formation and credit, and monetary behaviour. It is non-virtual: it deals with irrevocable exchange at prices that are disequilibrium ones from the point of view of the state of the entire set of markets, and with the non-virtual dynamics of production, consumption, saving and investment. Those irrevocable economic activities occur during the course of the equilibrating process and are part of it (1889, pp. 235, 280). The sub-models included in the comprehensive model, such as the models of consumer demand, of the firm, of the entrepreneur, of exchange, of production, and so on, will sometimes be called theories, because Walras had reference to the behaviour of the real economy rather than purely hypothetical schemes. Each major sub-model has four parts: structure, equilibration, equilibrium conditions and comparative statics.

Regarding the structure of each market, Walras assumed that preferences, the number of economically active individuals, the amounts of natural resources and technology are constant. He identified consumers, workers, landlords, capitalists and entrepreneurs, their economic characteristics, and their objectives and how they try to achieve them. He specified the types of goods that are traded, the institutional features and rules of the market, and the individual and market supply and demand functions for goods (material and immaterial).

Regarding the dynamic equilibrating processes by which the markets undergo adjustments when in disequilibrium, Walras called them 'tatonnements', which means 'gropings', to emphasize that the equilibrium magnitudes of prices and quantities are not known by the participants during the disequilibrium phase but are found by repeated trial and error experiments. Walras considered the exposition of tatonnement to be 'the object and proper goal of pure economics' because he believed that the real economy is stable (Walras to Bortkiewicz, 17 October 1889, L 927; Walras to Charles Gide, 3 November 1889, L 933). Walras gave a verbal demonstration of the stability of his model. He recognized that the dynamic functioning of markets depends on the economic agents, institutions and conditions identified in the first part of each model, and in order to portray the disequilibrium behaviour that he perceived in the real economy he accordingly discussed the activities and interactions among diverse economic agents in trade and production, the generation and elimination of profits and losses, the operation of the stock market and many other details of behaviour drawn from economic life. Most of his presentation of the model is concerned with its stability, that is, its behaviour in disequilibrium. Thus the allegation, perpetuated by generations of commentators (for example, Jaffé 1971, p. 281, 1981, pp. 252–61), that Walras devoted his attention almost exclusively to the conditions of static equilibrium in an abstract model devoid of institutional detail, economic facts and dynamic behaviour is a misrepresentation of his work.

Walras was partially responsible for that misrepresentation, because in 1900 he referred to his general equilibrium model as 'static' without qualification, and contrasted it with what he called 'the dynamic point of view', by which he sometimes meant the view taken in considering economic growth (Walras 1954, p. 318). On the other hand, he also stated on many occasions that a dynamic theory is contained in his mature comprehensive model, and his usage will be followed in this article. The 'static theory of exchange', he wrote, 'may be defined as the exposition of the equilibrium formula'. The 'dynamic theory', in contrast, which Walras claimed to have been the first to explore, is 'the demonstration of the attainment of that equilibrium through the play of the raising and lowering of prices' (1895, in 1965, vol. 2, p. 630). Similarly, in responding to Irving Fisher's criticism that he had not considered time, Walras pointed out that that was true only of his exposition of the conditions of static equilibrium, and that he gave a dynamic treatment of production in lesson 20 (1889) of the *Eléments* (Walras to Fisher, 28 July 1892, L 1064).

Theory of Exchange

Walras was concerned in this theory with the determination of the equilibrium prices of goods and the quantities of goods exchanged. Setting forth the structure of exchange markets, he assumed that the preferences of the traders and the aggregate amounts of the goods they hold in each market are given. He first assumed that goods (including services) are exchanged directly for each other and then that they are exchanged for money. The participants include brokers, professional traders, retailers, wholesalers, the owners of the factors of production in their capacities as demanders of consumer goods and capital goods properly speaking, and entrepreneurs, who supply and demand goods. The supply and demand functions are reciprocally related (Walras 1954, pp. 96–7). Given a trader's demand curve for A, its price times the related number of units he wants to buy is his supply of B expressed as a function of the price of A in terms of B. Observing what happens to the areas of the rectangles under the demand curve for A as its price rises, Walras deduced that the quantity supplied of B initially rises and then falls. In the same way, a trader's supply of A can be derived from his demand for B. Walras summed the individual demand and supply curves respectively in the market for A to obtain the market curves, and similarly for B. It will be seen that he adapted and extended this analysis of the dependence of the supply of one good upon the demand for another when he took up the question of multi-good exchange. Walras also assumed that in each market the rule is enforced that disequilibrium transactions are not allowed (1880a, p. 461; 1880b, p. 78; 1954, p. 85). He described that as being true of the nineteenth-century Paris bourse, but in fact disequilibrium transactions occurred there most of the time (Walker 2000, 2001), in recognition of which he allowed late in his career that his description is in actuality 'a hypothesis that no scientific spirit would hesitate to concede to the theoretician' (Walras 1895, in 1965, vol. 2, p. 630).

To explain demand and infuse his early model of exchange (1869–70) with purposive action, Walras developed a theory of preferences shortly before 1872 in which he assumed that traders want to maximize utility, that utilities are independent and additive, and that the marginal utility of a good is a decreasing function of the quantity acquired or consumed. Nevertheless, he was floundering in his attempts to relate utility to market behaviour, so he appealed for help to Antoine Paul Piccard, a professor of industrial mechanics at the Academy of Lausanne, who responded in 1872 by developing a model of utility maximization and deriving the individual demand function within it (1965, vol. 1, pp. 308–11), thus meriting a part of the credit that has previously been given exclusively to Walras for that achievement. Everything then fell into place for Walras, and he proceeded to develop the view of economizing and maximizing behaviour that he imprinted on Continental neoclassical economics. He extended the technique shown in Piccard's model, making utility maximization the driving force in each of his models, and obtaining the equilibrium conditions of the participants in a multi-good system (1954, lesson 12).

The dynamic behaviour of Walras's exchange model is a tatonnement process in the sense that the path of the price in each market is the unplanned outcome of market forces. The process depends upon human nature (see Walker 2006a, pp. 114–39) and on the rules, institutions and conditions devised and enforced by market authorities and by government (Walras 1880a, b; 1895 in 1965, vol. 2, p. 632; and see 1954, p. 474). A price is initially cried at random (1877b, p. 127; 1954, p. 169) by any of the traders, and the suppliers and demanders subsequently follow the Walrasian pricing rule: that is, they change the price in the same direction as the sign of the market excess demand for the good. Suppliers call out progressively lower prices if it is negative, and demanders call out progressively higher prices if it is positive. Preferences are constant, and the rule against disequilibrium transactions ensures that the asset distribution remains unchanged during the equilibrating process. Therefore, the initial supply and demand functions and, consequently, the particular-equilibrium price on any given day in the temporarily isolated market, are not affected by the disequilibrium behaviour of the traders. That price equates the supply and demand quantities; it is quoted sooner or later, and the equilibrium amounts of the good are exchanged (1954, p. 106, lessons 6, 9).

Markets are not isolated, however, so Walras introduced the central feature of his contribution to economic science, namely an account, in his theory of exchange and in the other parts of his mature comprehensive model, of the interrelationships among the markets for different goods (including services). If a trader has a good that he wants to trade for several others, the amount that he offers in any market is related to the amounts that he offers in the other markets, so the amount that he wishes to purchase or sell of any good is seen to be a function not only of his preferences, his income and the price of that good but also of the prices of other goods. Consequently, the market supply and demand quantities and the price in any market are dependent in part upon the prices in other markets (1954, lesson 12).

Moreover, Walras explained that the sum of the values of a trader's quantities demanded must equal the sum of the values of his quantities supplied. That relation is one way of stating the individual budget equation, and it is a version, on the individual level (1954, p. 165), of what has come to be known as Walras's Law, a fundamental statement of the way that markets are interrelated. Walras was able to identify the law in part by reasoning an individual cannot demand any commodity without offering in return commodities (or money) having the same total value, so, if some of his excess demands are positive, others of them must be negative, and in part because of the device of the numeraire. The latter, a good in terms of which the values of all goods are expressed (1954, p. 161), made clear to him, as it had to Isnard, that there is exactly the right number of excess demands: in a system with n goods, there are only n – 1 independent market equations involving n – 1 price ratios, but also only n – 1 unknowns, inasmuch as the price of the numeraire, the nth good, in terms of itself is unity (1954, pp. 161–2, 241).

With reference to the market level in multicommodity exchange, Walras affirmed that the sum of the positive or negative market excess demand quantities for each good multiplied by its price is zero (1954, p. 170), and he stated a version of Walras's Law for the market excess demand quantities of productive services (1954, p. 248). In a Walrasian equilibrium, supply equals demand for every good, so each excess demand quantity is zero. Each excess demand quantity, multiplied by the price of the good, must therefore be zero, so the sum of the excess demands each multiplied by the price must be zero. In the case of an individual, Walras stated only, regarding the variables, that 'there will be between them all' the relationship that indicates that their sum is equal to zero, without addressing explicitly whether the equation is true in disequilibrium as well as in equilibrium (1889, p. 143, which differs from Walras 1954, p. 165). In the case of multicommodity exchange, he probably implied that it is always true even though in disequilibrium the market supply and demand quantities of every commodity are not simultaneously equal (1954, pp. 169–70). In the productive services formulation, he indicated explicitly in the *Eléments* that it

is true only in equilibrium; the functioning of the market mechanism is necessary when the economy is not in equilibrium, he stated, to drive the excess demands to zero and thus solve the equation (1954, pp. 248–9; 1889, pp. 242–6). He later declared, however, that it holds in both disequilibrium as well as in equilibrium (1898, pp. 277–8; Walker 2006a, pp. 152–4). His implicit reasoning appears to be that the *sum* of the excess demand quantities, each multiplied by its price, is zero in disequilibrium even though some or all excess demand quantities are not zero, thus implying that the law is an identity.

Walras asserted regarding his mature comprehensive model (and hence regarding its submodels) that equilibrium exists, on the grounds that the number of independent equations equals the number of unknowns (prices and quantities). He was, of course, mistaken in that belief. Wilhelm Lexis had pointed out in 1881 that Walras's equations might nevertheless not have real positive solutions or any solution at all (see Walras 1965, vol. 1, p. 747), a fact that only since the late 1920s and early 1930s became well-known (see Weintraub 1983; Van Daal et al. 1985).

The interdependence of markets, Walras explained, gives rise to the major problem of general equilibrium analysis, which is the question of the stability of the model, implicitly containing the question of the existence of equilibrium. Will a system of freely competitive markets that is initially in disequilibrium converge to a position of equilibrium? After any market reaches temporary equilibrium through the exchange of the equalized market supply and demand quantities, the traders note what has happened to the prices in other markets. Their reaction is manifested in a shift of the market demand curve, which puts the market once more into disequilibrium and initiates another series of quoted prices leading to a new market-day equilibrium. Will its readjustment aid or impede the equilibrating process taking place in other markets? Does the series of market-day prices in the set of markets move closer to an equilibrium of the entire system or further from it? Walras claimed that he had shown that the answer to those questions is that the market system converges to general equilibrium as a result of the ways that markets are interrelated and of the operation of the Walrasian pricing rule in each market (Walras 1954, pp. 172, 179–80).

Walras then specified the conditions that prevail in the static equilibrium of exchange of a multi-market system. The ratio of the raretés, or marginal utilities, of any two goods is equal to the ratio of their prices, and the price of any good in terms of another good is equal to the ratio of the prices of those two goods in terms of any third good (1954, p. 157). Those conditions are satisfied when the quantities supplied and demanded of each good are equal (1954, p. 172).

Finally, Walras briefly examined some features of the comparative statics of the exchange model (1954, pp. 147–9). He shifted the utility curves for a good and determined that its equilibrium price changes in the same direction as the shift in the curves. He then successively increased and decreased the traders' endowments of a good and determined that its equilibrium price successively decreases and increases.

Theory of Production

In his model of production, Walras was concerned with the determination of the equilibrium prices of productive services and the equilibrium rates of output of the quantities of non-durable goods. Walras did not present this model directly and without modifications as part of his mature comprehensive model. The latter deals with durable goods as well as non-durables and therefore contains a wider model of production. Nevertheless, Walras carried over the processes of pricing and aspects of the tatonnement in the non-durables model into the comprehensive model.

Setting forth the structure of the production model, Walras first identified the markets for productive services, in which he assumed that the amounts of economic resources and therefore the maximum possible amounts of their services are given. The demanders of productive services are the entrepreneurs. Their ultimate aim is to maximize utility, which they achieve through maximizing profits. In their capacities as entrepreneurs, they combine productive services and materials

in proportions that are determined by what Walras called the technical coefficients of production. The coefficients, which he assumed to be fixed in much of his general equilibrium theorization, indicate the amount of each of the inputs that is used to make a unit of output. With fixed coefficients and given prices of the productive services, the average cost is constant as the firm's output varies. If any of those prices change, the average cost curve shifts in the same direction.

The suppliers of productive services are workers, who own personal faculties; landlords, who own natural resources; and capitalists, who own capital goods or provide capital funds (Walras 1877b, p. 218; 1954, pp. 214–15). Their aim is to maximize utility, which motivates them to offer services to the entrepreneurs in exchange for income.

Walras then identified the market for consumer goods (material and immaterial). These goods (in the production model) are consumed immediately after being produced; they are used only once and are used up in that process. The suppliers of them are the entrepreneurs. The demanders are the workers, landlords, and capitalists acting in their roles as consumers, motivated in their purchases by the desire to maximize utility. They pay for them with the incomes that they have been paid by the entrepreneurs. The only type of capital goods produced in the model is nondurables, that is, variable capital goods like raw materials. Those goods, like consumer goods, are used up in a single application as soon as they are purchased.

Of course, that is true of the services of all types of economic resources. In the model, there is no saving. The durable capital goods that are used in production do not depreciate or become obsolete, nor are they subject to accidents. There are no markets for them or for land.

The tatonnement in the production sub-model, and in the capital goods and monetary sub-models, and in the comprehensive model as a whole, is in considerable measure the outcome of the actions of entrepreneurs. Walras assumed that all resources are highly mobile. Entrepreneurs have good knowledge (but not perfect foresight) of the profitability or unprofitability of producing any particular good and accordingly enter or leave an industry. The tatonnement that occurs in the markets for inputs is a process of groping for the equilibrium amounts of resources employed in different industries. The entrepreneurs hire the factors of production, combining them in technologically determined proportions or experimenting to find optimum proportions if the coefficients are variable (1896a, pp. 490–1), and sell services and finished goods to consumers (1954, lesson 21, and pp. 426–7; Walker 1996, ch. 13). The entrepreneurs hire and use disequilibrium quantities of productive services during the tatonnement, and produce disequilibrium quantities of goods (1889, pp. 234–5, 240–1, 249–50). The payment that the entrepreneurs receive in disequilibrium for their entrepreneurial activity is profit, which Walras defined on a per unit basis as the price of output minus its average cost, with the latter including the wages of management. An entrepreneur may undertake the functions of other factors of production – he may also, for example, be a capitalist or a manager of the firm – and ordinarily he has to do so, but his multifaceted role as entrepreneur is a distinct one (1954, p. 222).

The tatonnements in the markets for productive services and for consumer goods are interrelated. If the consumers' demand for a good increases, the price is bid up in accordance with the Walrasian pricing rule. The quantities demanded and supplied become equal at a high price because the supply function does not initially change. The price of the product then exceeds its cost of production, so the entrepreneurs in the industry make profits. Attracted by the prospect of doing the same, other entrepreneurs enter it, and existing firms increase their output. As the market supply of output function changes so that more output would be offered at each possible price, the price in the exchange market for the good is lowered by the entrepreneurs in an effort to dispose of the entire flow of output. As the demands for productive services increase, their prices are bid up, which raises the average cost of production. Thus the price falls and the average cost rises (1954, p. 253). If demand for a good decreases, its price falls below the average cost of production and the entrepreneurs make losses. This leads

some of them to leave the industry and some of them to diminish the output of their firms. The prices of the productive services fall as the demand for them decreases, which lowers the average cost of production. As less of a good is offered, its price is forced up. Thus the average cost falls and the price rises (1954, p. 253). It will be noted that these are all non-virtual processes in the model. Walras concluded that the average cost of production and the price of the good become equal, whereupon equilibrium is reached and the tatonnement ends. It follows that in equilibrium the entrepreneur obtains neither profit nor loss (1877b, p. 232; 1954, p. 225).

Pursuing still further the question of the stability of the model, Walras reasoned that the interactions of traders in different markets aid the equilibrating process. The change in the output of a product, he argued, has an effect on its price that is unidirectional, whereas the unidirectional changes that it induces in the outputs of other products have only indirect repercussions on its price, and the latter more or less cancel each other because some of them change in one direction and some in another (1954, p. 246). In the non-virtual adjustment, 'The [resulting] system of new quantities of manufactured products and of new selling prices is thus closer to equilibrium than the old one; and we have only to continue the process of groping to approach still more closely to equilibrium' (1954, pp. 246–7). In other words, once again, ultimately the tatonnement leads to the simultaneous equalization of supply and demand in every market.

The equilibrium, Walras stressed, is the normal state of the market in the sense that it is the one to which the variables in disequilibrium perpetually and automatically tend in a regime of free competition (1954, p. 224). Since it contains the equilibrium of exchange (1954, p. 224), it is characterized by the conditions that the quantities supplied and demanded are simultaneously equal regarding each consumer good, each productive service, and each primary material. A stable circular flow is established in which the total cost equals the total revenue in each firm, the incomes received from the entrepreneurs by the owners of the factors of production equal the revenues earned by the firms, and those incomes are spent on consumer goods by the owners of the factors of production. Walras's theory of production therefore showed ways in which input and output markets for non-durable goods are linked together.

Sales at a disequilibrium price of the items that are produced do not alter the assets held by the participants because the quantities exchanged do not have an existence after the exchange; they are used up immediately. The equilibrium of the production model is therefore not path dependent. That is of no significance, however, because the model is a hypothetical special case. It does not take account of the production of capital goods or of consumer durables which occurs in the real economy and in Walras's mature comprehensive model.

Walras then considered variations in some of the parameters of the production model. If the marginal utility curves for a good shift up, he reasoned, its price in terms of the numeraire increases. If the marginal utility curves shift down, the opposite occurs. If the quantity of a product or service possessed by the holders changes, its price changes in the opposite direction (1954, p. 260).

Theory of Capital Formation and Credit

Walras referred to the three sources of services (labour, land, and capital) as different types of capital because they all endure through time and produce a flow of services, but in this article the unqualified word 'capital' or the term 'capital goods' will mean durable, man-made, inanimate instruments of production. Walras first examined the determination of the prices of land and personal faculties, as distinct from the prices of their services. The aggregate supply of land, a given condition, is perfectly inelastic, and its price is its gross income divided by the rate of net income (1954, pp. 270, 309). The number of workers is also a given condition. The price of a worker is equal to his gross income minus the cost of replacing and insuring him, divided by the rate of net income. Workers are not bought and sold, however, so their prices are virtual (1954, pp. 271, 311).

Walras's theory of capital is concerned with the determination of the amounts and prices of capital

goods, as distinct from their services, and the determination of their rate of net income. Capital goods are specific items of real capital; capital funds, raised by the sale of shares on the bourse, constitute fluid and mobile purchasing power which can be used to acquire economic resources to construct different kinds of physical capital (1954, pp. 270, 311). Capital is formed by capitalists saving money and, most commonly, lending it to entrepreneurs (1954, p. 270). The net saving of the capitalists as a group equals aggregate income minus aggregate consumption, minus the depreciation and insurance costs of capital goods. The entrepreneurs purchase or rent capital goods, earn revenue from their use, and repay any sums they have borrowed (1954, p. 290, §§ 190, 235). Walras's identification of that process explains why he inserted the word 'credit' (1954, p. 270) into the name of his capital-goods model, but he did not develop a general theory of credit within it, because he did not introduce loans made by banks. Some capitalists prefer to own capital goods, so Walras assumed occasionally that they acquire them directly in physical form (1954, p. 289), and assumed frequently that they acquire them through buying stock certificates (for example, 1954, p. 289). In each case, the physical capital is used by entrepreneurs, so 'the demand for new capital goods comes from entrepreneurs who manufacture products and not from capitalists who create savings' (1954, p. 270). The entrepreneurs purchase the particular kinds of capital goods that are profitable, with the result that the kinds that are produced and used reflect the structure of demands for consumer goods (1954, pp. 225, 303).

Describing the tatonnement regarding both non-durable products and durable goods, that is, summing up the non-virtual character of the tatonnement in the mature comprehensive model, Walras explained that

> After a certain rate of net income and certain prices of services have been cried and after certain quantities of products and new capital goods have been manufactured, if this rate, these prices and these quantities do not satisfy the conditions of general equilibrium, it will be necessary not only to cry a new rate and new prices, but also to manufacture revised quantities of products and new capital goods. (1889, p. 280; 1954, p. 282; and see 1889, pp. 284–94; 1954, § 258, pp. 293–4)

One aspect of the tatonnement takes place in the stock market and another in the course of the production of capital goods. In the stock market, which is the market for new capital goods, each capitalist attempts to maximize utility by saving and acquiring more stocks that have relatively high yields and less of those with lower yields (1954, p. 289), with the result that the total value of new capital goods and the excess of income over consumption both move in the same direction as all prices. It follows, Walras maintained, that the tendency of the change in prices to destroy the equality between the total value of new capital goods and the excess of income over consumption is weaker than the tendency of the change in the rate of net income to bring the total value of new capital goods and the excess of income over consumption into equality with each other. Moreover, 'in these conditions, it is evident that the price and the average cost of the capital good (K) will have been little altered from equality by the increases in the quantities produced of the capital goods (K′), (K″),. . .' (1889, pp. 292–3). 'Thus the system involving the new rate of net income and the new prices will be closer to equilibrium than the old system; and it is only necessary to continue the process of groping for the system to move still more closely to equilibrium' (1954, p. 288).

In the course of the production of capital goods proper, entrepreneurs acquire more capital goods that yield relatively high returns, and diminish their use of capital goods that yield lower returns. As a consequence, the rate of net income from each capital good proper tends toward the same value (1954, p. 309). If profits are being made from the production of capital goods in an industry, new entrepreneurs enter it, and those already in it increase their rate of production. That drives up the prices of productive services, which causes the average cost to rise towards equality with the price of the capital good. If losses are incurred, entrepreneurs diminish production. That drives down the prices of productive services, which causes the average cost to fall towards equality with the price of the capital good (1889, p. 293). It

is 'probable', or, 'it is to be presumed', Walras maintained, that the effects of changes in the output of a new capital good that tend to cause equality between its average cost and its price will be stronger than the contrary effect of interrelated changes in the output of other capital goods, so the process converges to equilibrium (1954, p. 293). Referring again to the non-virtual character of the tatonnement, Walras explained that 'The new system of revised quantities manufactured and of revised costs of production and selling prices of new capital goods is thus nearer equilibrium than the original system', and it is only necessary to continue the tatonnement to approach it more and more closely (1889, p. 293; 1954, p. 293).

The tatonnement in the mature comprehensive model involving both nondurable and durable goods represents and explains what happens in the real economy:

> Now, this tatonnement is precisely that which occurs by its own forces on the market for products [non-durables], under the regime of free competition, while the entrepreneurs who produce new capital goods, just like the entrepreneurs who produce products, increase their production or diminish it according to whether profits or losses are made. (1889, p. 294; 1954, p. 294)

Walras expressed the equilibrium conditions in the formation of new capital goods in the lengthy analysis that he called the theorem of the maximum utility of new capital goods, which he described as crowning and confirming his entire theoretical system (Walras to H.S. Foxwell, 16 December 1888, L 859; Walker 1984b). Although he assumed initially that new capital goods do not require amortization or insurance, Walras then made the realistic assumption that capital goods wear out and are subject to accidents. The rate of net income generated by a particular capital good is then given by the gross income it earns minus amortization and insurance costs, divided by the price of the capital good. In equilibrium each trader maximizes utility by holding the quantities of capital goods that make the ratio of the marginal utility of each capital good to its price equal for all his capital goods. Because of the adjustment of yields and capital good prices, the rate of net income derived from every type of capital good is the same (1954, p. 281). There is a single price for each type of capital good, equal to its average cost (1954, p. 280). Furthermore, the price of any given type of well-maintained old capital goods is equal to the price of the same kind of new capital goods, so the equilibrium prices of all capital goods 'are equal to the ratios of their net incomes to the rate of net income' (1954, p. 309). The latter is the rate of interest. Its equilibrium value equates aggregate saving and investment (1954, p. 276).

Walras believed that through this analysis he had seen behind the veil of money or numeraire and discovered the real determinants of that rate. It is manifested in the banking system, he argued, but it is determined in the stock market. It is the ratio of net profit to the price of a share of stock, which in equilibrium equals the common ratio of the net price of capital services to the price of the good that yields them (1954, p. 290).

Walras then turned to the comparative statics of the capital goods market. If the price that has to be paid for the services of a capital good increases or decreases as a result of a parametric change, the price of the capital good itself increases or decreases. Its price also varies inversely with the rate of depreciation and with the rate of the insurance premium. If the rate of net income changes, the prices of all capital goods change in the same direction (1954, pp. 309–12).

Theory of Money and Circulating Capital

Walras wanted to reduce the monetary mechanism 'to its essential elements' (1889, p. 379). He therefore carefully specified the structural and behavioural features of his mature fiat money model, drawing upon his direct experience as a young man with real financial matters and his extensive knowledge, accumulated throughout his career, of empirical monetary arrangements and problems. He explained that fiat money, like pieces of paper with an engraved image, has no utility of its own. Economic agents hold it because it enables them to purchase goods that have utility (1889, p. 378). Entrepreneurs and some consumers have net demands for cash balances because of the non-synchronization of payments

and receipts (1886, pp. 40–4). The size of a consumer's desired cash balance depends upon desires to consume and save, which depend upon his character and habits, income, the value of the commodities he wants to buy, and the rate of interest (1889, pp. 377, 268–71). The size of an entrepreneur's desired cash balance depends upon the nature and state of his business, and on his character, habits, and the rate of interest (1889, p. 377). These are desired real balances because they represent the demand for a specific bundle of goods. The aggregate real desired cash balance is the nominal one divided by the price level, which is the inverse of the price of money. That aggregate is the demand for future consumer and capital goods (1889, pp. 378–9).

Walras incorporated a market for loans into his model. First, there are the demanders. They are consumers, entrepreneurs and savers, who go to the market and borrow. The first two groups buy the goods and services they need. A curious temporary assumption is that savers, the third group, borrow to obtain the funds they lend. Thus 'we have the daily demand for money which is exercised on the market for money capital' (1889, p. 381). What is happening behind the veil of money, Walras explained, is that the consumers and entrepreneurs are actually borrowing the fixed and circulating capital on which they spend the money, and the interest borrowers pay is generated by the fixed and circulating capital that the borrowed money finances. Second, there are the suppliers of funds in the loan market. In one group of them are entrepreneurs who have receipts from the previous day from sales to consumers and from sales of new fixed and circulating capital good to other entrepreneurs, and in the other group are landlords, workers and capitalists who have receipts from the previous day from sales of services (1889, p. 381)

Walras then specified how the mechanism of free competition operates in regard to monetary circulation in disequilibrium of the money market. It will be noted that the entire behaviour relating to money in the mature comprehensive model is nonvirtual. Mainly explicitly, he indicated the tatonnement that was presented later by the Cambridge cash balances theorists. He created the temporal framework for the period analysis in his model by assuming first that production and consumption 'extend over all the moments of the entire year' (1889, p. 316). Workers, capital goods, and money are used up and are simultaneously reproduced and replaced. He then assumed that markets function every day and that different types of related behaviour occur sequentially on a series of days (1889, pp. 381–3). Thus he developed a continuous-production and periodic-market model.

A change, say a decrease, in the quantity of money causes disequilibrium. The Walrasian period analysis then indicates that the process of equilibration takes three 'days'. On the day that the decrease occurs, the old rate of interest still rules and the quantity of cash balances demanded exceeds the quantity supplied. The immediate result is that the rate of interest increases. On 'the next day on the market', a temporary equilibrium is reached, 'at a new higher rate of interest at which the desired cash balance would be reduced' (1889, p. 383). During that day, the prices of all goods fall proportionately to the decrease in the quantity of money and the aggregate *nominal* desired cash balance remains lower, but the real balance is 'able to become what it was before' as a result of the fall of prices. Then, 'on the day after that', the third day, permanent equilibrium is attained, the rate of interest falling back to its old level, which is equality with the rate of net income from capital goods (1889, p. 383). The same sequence occurs if the aggregate desired cash balance function increases. If the parameters change in the opposite direction, the opposite sequence of adjustments occurs. There are many more details, situations and variations of assumptions that Walras considered regarding the model. He was able to sum up what he had done in this way: if the quantity of money increases or the desired cash balance decreases, prices rise proportionately, and the reverse. 'This law extends to money the principle by virtue of which value increases with utility and decreases with the quantity' and therefore integrates the determination of the value of money into his mature comprehensive model and therefore into his general theory of value (1889, p. 383).

When there is monetary equilibrium, that model is in equilibrium in all respects. Walras summarized its aspects in the following way. There is 'the equilibrium of exchange' in which prices are proportional to marginal utilities and consumer satisfaction is maximized; there is 'the equilibrium of production' with prices equal to average costs and zero profits; there is 'the equilibrium of capital formation' with prices of land, human faculties and capital goods proper proportional to the prices of their services; and finally there is 'the equilibrium of circulation, in view of the fact that the exchangers would have the cash balances that they desire at the announced rate of interest' (1889, p. 379).

Walras therefore made significant theoretical innovations in his mature theory of money. In it, he raised basic questions about the nature of a true money, its role, the valuation of money, its place in preference functions and hence in demand and supply functions, the sequence of adjustments that occur after a change in its quantity, and the impact upon equilibrium prices and the rate of interest. His modelling of cash balance behaviour and dynamic period analysis anticipated some of the theoretical techniques used during the 1920s and 1930s by J.M. Keynes,D.H. Robertson, and J.R. Hicks, none of whom acknowledged his contributions. That was probably because the insightful analysis and potentially fruitful constructions that Walras put into the mature money model are in the forgotten 1889 edition.

Economic Growth

Walras did not develop a complete model of economic growth, but he examined some aspects of the topic in connection with his mature comprehensive model. One was endogenous variations in its parameters. He was led to speculate about that subject by the consideration that the equilibrium conditions identified in the mature comprehensive model are never reached in the real economy because tatonnement takes time, and consequently parameters such as preferences and the amount of labour change before equilibrium is reached (1954, p. 380). In order to take account of this situation, 'we shall now suppose that the annual production and consumption, which we had hitherto represented as a constant magnitude for every moment of the year under consideration, change from instant to instant along with the basic data of the problem' (1954, p. 380). Services and goods are used up and are produced. Net new capital goods come into existence and are put to use, and circulating capital is borrowed by entrepreneurs from the capitalists in the form of money loans (1954, p. 379). Walras did not analyse in any further detail that moving equilibrium or 'continuous market' economy.

A second aspect of growth that he examined was 'the laws of the variation of prices in a progressive economy' (1954, p. 382), that is, some of the features of alternatives paths of economic growth. For this task he first defined economic progress as the substitution of capital services in place of land services in given production functions (1954, p. 383). The substitution implies variable coefficients of production, and to introduce these Walras used the theory of marginal productivity. He did not claim to have originated that theory although he anticipated some of its features. In fact, Hermann Amstein, a mathematician at Lausanne, worked out its major principles in 1877 (Amstein to Walras, 6 January 1877, L 364; translated in Jaffé 1983, pp. 205–6). Walras did not understand or use Amstein's work, however, and the major credit for the theory of marginal productivity that first appeared in the *Eléments* in 1896 (Appendix III) must be given to Enrico Barone (1895).

Walras defined technical progress as changes in production functions, including the introduction of entirely new processes, but he did not analyse it, beyond concluding that it contributes, along with economic progress, to ensuring that output increases without limit in a progressive economy (Walras 1954, p. 387). He also discussed, in a highly general way, how the prices of products and services vary with different amounts of capital and sizes of the population (1954, pp. 389–91). His principal conclusion was that the rate of net income falls as the stock of capital grows, the proximate causes of the process being rising rents and falling prices of capital services.

Walras was well aware that capital accumulation means economic growth and requires a

different characterization of equilibrium, noting that 'In order to have a supply, a demand, and prices of capital goods, it is necessary to substitute for the conception of a stationary economic state that of a progressive economic state' (1889, p. 264). His way of dealing with this problem in 1889 was to assume that new capital goods are not used until what he represented to be the end of the tatonnement, thinking that would preserve the static equilibrium. That did not, however, remedy the problem, as will now be shown.

Walras attempted to give a mathematical proof of the stability of the tatonnement of the mature comprehensive model, spread throughout the pages of his treatise. He believed that he showed that the model is stable by working with his system of static equations of general equilibrium. He posited a disequilibrium state and then varied the prices in the equations in accordance with the Walrasian pricing rule, and claimed that equilibrium is the result of the tatonnement. That claim is logically flawed, for two reasons.

The first is that the tatonnement in the mature comprehensive model, unlike the model of the production of non-durable goods and services, is path dependent even though the new capital goods are not used during any of the phases of adjustment before equilibrium is ostensibly reached. Positive net investment has the result that individual holdings of capital goods and many of the other nominal parameters and all the variables of the model are altered. Each different disequilibrium rate of production and sales of capital goods occurring within each phase of the tatonnement changes their prices and average costs, profits, and the rate of net income. Consequently, Walras's attempted proof was not rigorous and could not have been valid, given the static equations that he used. They have the endowments of assets in the initial disequilibrium as parameters. Their solutions therefore depend on those constants. The model, however, is not a virtual system, so the individual holdings of assets and their total amounts vary during the course of the tatonnement. The result is that the variables of the model do not converge to the solutions of Walras's static equation system. Any equilibrium to which the prices and quantities converge cannot be the one indicated by his equations because they do not describe his model.

The second reason is that the 'equilibrium' that Walras asserted exists at the end of the tatonnement is factitious and cannot materialize, even if there were no problem of path dependency. The supposed equilibrium could exist only transitorily while the model is held in a state of arbitrarily suspended animation by the postulate that the additions to the capital stock are not used – a *deus ex machina* that interrupts the incomplete workings of its endogenous processes. The instant that Walras removed the postulate, that is, the instant that the net new capital goods are put into use, the 'equilibrium' is ruptured, and through dynamic processes many of the nominal parameters and all the variables of the model change, in the way just indicated in the discussion of the consequences of the use of net new capital goods. If they continue to be produced, as Walras assumed, the system follows a path of growth. The equilibrium path and any stationary equilibrium that the system may eventually reach is quite unlike the solutions to the static equations of general equilibrium that he presented in the 1889 edition of the *Eléments* and subsequently.

Walras's Last Theoretical Work

The Written Pledges Sketch

In 1899 Walras changed his work in two major ways, and put the changes into the *Eléments* in 1900. One was to devise a new model of money and circulating capital (see below). The other was to try to construct a virtual model. The motive for the latter was that Walras had come to realize by 1899 that his equation system is compatible only with such a model. Instead of trying to develop a different equation system, however, one that would represent the non-virtual mature comprehensive model, he chose to abandon the latter, to retain his static equations, and to try to construct a virtual model that would serve as their foundation and justification. In the 1900 revision, he eliminated much of each of two forceful and lengthy statement of the non-virtual tatonnement (Walras 1889, pp. 234–5, 280), which consequently

appear in Jaffé's translation only in very abbreviated form (1954, pp. 242, 282). He retained, however, crucial parts of those statements and retained elsewhere throughout the revision most of the other language describing the non-virtual behaviour of the economy and of the mature comprehensive model. That explains why the reference 'Walras 1954' can be cited to document many the features of the 1889 mature comprehensive model. It is also one of the principal causes of the 1900 and 1926 editions (and therefore Jaffé's translation of the latter) being a chaotic mixture of incompatible language and sub-models.

To construct a virtual model, Walras conceived the device of written pledges (*engagements écrits,* as they were and are called in the Paris bourse). He asserted that the model has three phases, made identifiable, he believed, 'by means of the hypothesis of written pledges'. First, there is 'the phase of preliminary gropings towards the establishment of equilibrium in principle', the purely virtual phase (1954, p. 319). When a price is cried in any market, suppliers of goods and services write out the amounts that they pledge to produce and sell at that price, but only if it turns out to be the equilibrium value, that is, the one that is part of the set of prices that equates the market supply and demand quantities simultaneously in every market (1899, p. 103; 1900, pp. 215, 260; 1954, pp. 242, 282). Thus Walras adopted the device in order to eliminate changes in the holdings of assets before the entire system of markets has reached equilibrium, changes which would otherwise occur as a result of trade occurring in a market when it reaches market-day equilibrium while other markets are still in disequilibrium, or as a result of disequilibrium production, which changes the aggregate amounts of goods held before general equilibrium has been reached.

In the first phase, entrepreneurs are supposed to plan to move from unprofitable to profitable industries and to plan to create firms or to expand or contract their existing firms, and to predict accurately the financial results of their plans, without actually moving or creating or hiring or spending or producing at all. Owners of productive services are imagined to offer their services repeatedly at disequilibrium prices without actually earning any income or consuming any goods or services. The entire system of interrelated markets is imagined to go through complex costless processes of information acquisition, price changes and changes in the supply quantities that are pledged, all without anyone being allowed to agree to a single actual transaction or undertake any production or consumption, until the equilibrium set of prices has been found.

It is symptomatic of Walras's significantly diminished ability to concentrate and pursue lengthy chains of reasoning after 1898 that his words 'the hypothesis of written pledges', although followed by dozens of pages of modelling and theorization, including his immediately following account of the three phases which he introduced into the fourth edition of the *Eléments* at the same time as writing those words, is the last sentence in which Walras mentioned them in any of his writings. His references to written pledges had become fewer and fewer in the successive pages of the 1900 revision. Finally he either forgot about them or decided they were not an idea worth pursuing any further, abandoned his attempts to change the older constructions in his treatise to accord with their use, and introduced new submodels in which they are not used.

Second, there is 'the static phase in which equilibrium is effectively established *ab ovo* as regards the quantity of productive services and products made available during the period considered, under the stipulated conditions, and without any changes in the data of the problem' (1954, p. 319). This means that the economy 'remains [for the time being] *static* because of the fact that the new capital goods play no part in the economy until later in a period subsequent to the one under consideration' (1954, p. 283). In this postulated static equilibrium, services and nondurable consumer commodities are produced, sold and used. Walras was asserting that the result of the tatonnement in the sketch is that the market supply and demand quantities are equated in every market simultaneously, whereupon the non-virtual activities of exchange, production, consumption, saving and investment take place. He therefore believed that none of the parameters ('the data of the problem') of his system of equations of

general equilibrium undergoes endogenously induced changes during the tatonnement. He believed that a static equilibrium is consequently the one given by the solutions to his static equation system, and that his new version of tatonnement converges to that equilibrium for the same general reasons as he had adduced in 1889.

Third, continuing to write as though the sketch were a complete functioning system, Walras indicated that it undergoes 'a dynamic phase in which equilibrium is constantly being disturbed by changes in the data and is constantly being reestablished' (1954, p. 319; 1900, p. 302). The new capital goods, both fixed and circulating, Walras wrote, 'are made available during the second phase' but 'are not put to use until the third phase'. When they are used, however, 'the first change in the data of our problem' occurs (1954, p. 319). The result of changes in the data is that the *'fixed* equilibrium will then be transformed into a *variable* or *moving* equilibrium, which re-establishes itself automatically as soon as it is disturbed'(1954, p. 318). The use of the new capital goods generates economic growth.

Of course, Walras *asserted* that the three phases and all the behaviours and outcomes that he wanted the sketch to have, such as the equalization of supply and demand, do in fact occur in it, but in actuality the sketch does not support his claims. Those accounts are not descriptions of the sketch. They are simply postulates; they cannot be deduced from its structure and procedures, which is why Walras's work on virtuality is properly described as a sketch rather than a model. He made the mistake of assuming that potential demanders, whether consumers or entrepreneurs, do not write pledges to buy, so they have no way of making their desired demands known. For that reason alone (although there are many others; Walker 1996, section I), the sketch cannot function. There are no individual or market demand functions, Walrasian pricing, transactions or production. Equilibrium does not exist because the number of unknowns (prices, the quantities of goods and services) exceeds the number of independent equations (the market supply functions). Moreover, the device of written pledges is so inherently flawed that there are no conceivable revisions that can transform it into a functioning system (Walker 1996, section I). Finally, it is evident that the sketch has no explanatory value in reference to the real economy.

Despite the sketch's shortcomings, its aim of excluding irrevocable disequilibrium behaviour from a general equilibrium model, achieved simply as a postulate, nevertheless became a central and indispensable part of Walras's intellectual legacy for certain of his successors, including Gustav Cassel, Abraham Wald, John von Neumann, Kenneth Arrow, Frank Hahn and Gérard Debreu (Walker 2006a, pp. 288–312). It is a pity that they adopted his virtual notion and were unaware of or disregarded his robust and more realistic mature comprehensive model, for through the development of many of its characteristics lies the way to a more useful general equilibrium theory, as recent research has shown (Walker 2006a, pp. 334–6; 2006b).

Commodity E in the Last Model of the Production and Pricing of Capital Goods

In 1900 Walras invented a fictional good (E) constituted of perpetual annuity shares, another example of the deterioration of the quality of his modelling after 1899. It was by means of that concept that he chose to deal with a positive, zero or negative excess of aggregate income over aggregate consumption in his capital-goods model. It appears that he wished to express aggregate savings as a single homogenous quantity – the demand for E. Each perpetual share pays one unit of numeraire per year, and its price, determined by supply and demand, is the reciprocal of the rate of interest (Walras 1954, pp. 274–6, 308–9). If people want an additional amount of interest income, they provide savings through purchasing new perpetual shares. The numeraire-capital that capitalists pay for units of E is used by entrepreneurs to buy productive services and materials that are transformed into new capital goods).Walras viewed the capital-goods markets as reaching equilibrium through adjustments of the goods' rates of return (1954, pp. 275–6, 308).

Walras's device of the E commodity, which he frankly described as imaginary (1954, p. 274), has

not been adopted by economists who succeeded him, because of its remoteness from the realities of capital accumulation and the distortions that it creates in a model of that process. He did not realize that it is incompatible with the written pledges procedure. If the latter is assumed to occur, the capitalists have no way of expressing their demand for E. Moreover, as it happens, Walras retained his references to the purchasing of stock certificates and private and government bonds which appear in the mature capital goods model (for example, 1954, §§ 255, 270), and he introduced short-term loans into his new money model, although the incompatibility of E with those financial instruments increases the incoherence of the last two editions of the *Eléments*. His treatment of saving, investment and the capital goods market in the mature comprehensive model is superior to his final thoughts on the subject.

Walras's Last Model of Money and Circulating Capital

In his revision of the *Eléments* in 1900, Walras stated that he wanted to design the structure of this model on 'exactly the same terms and in precisely the same way' as in the 1900 models of exchange, production and capital formation (1900, p. 42). He did not in fact do that because he mentioned written pledges only twice in the first half of the exposition of the model and not at all in the second half of it. In fact, he constructed the new model of money and real circulating capital before he thought of written pledges. He first presented it in an article in 1899. After the last page of the article, he added an afterthought, a note of 37 lines introducing the device of written pledges (1899, p. 103). He subsequently inserted most of the note into the text of the article, and inserted the result almost verbatim into the 1900 edition of the *Eléments* (1954, lessons 29, 30), completely eliminating the treatment of money that he had presented in the mature comprehensive model.

In contrast with his mature comprehensive model, Walras described the functions of money and the formation of money prices in his last model of general equilibrium on the assumption that there is no uncertainty in equilibrium, and consequently that the dates and monetary value of future purchases and sales are known (1954, pp. 317–18). Money is one type of circulating capital, he explained; the other is circulating physical capital. Replacing his formulation of an equation of exchange that had anticipated Irving Fisher's (1877b, pp. 180–81), Walras asserted that circulating physical capital yields utility from its 'service of availability' -that is, by being readily available -and money provides, by proxy, the same service of availability as the good that it is destined to purchase and yields the same utility as that service. All economic agents try to hold utility-maximizing amounts of money and circulating physical capital (1954, pp. 320–1). The latter, held by consumers and entrepreneurs, is acquired with money, so the essential concern of Walras's model of circulating capital reduces to the question of the demand for and supply of money and its price.

Savers make some of their balances available as loans through buying or perpetual annuity shares (1954, pp. 318–20). The aggregate gross supply of money is the total stock issued by the monetary authority in the case of a fiat money economy, and is the amount of circulating coin in the case of a commodity-money economy (1954, pp. 321–4). The price of money, the numeraire, is unity and the price of its service is the rate of interest (1954, pp. 320, 327). Given the flows of receipts and purchases, the individual gross and net demand for cash balances and the individual net supply of them are functions of the rate of interest. The sum of the individual net demands for money is the aggregate demand function, and the sum of the individual net supplies of money is the aggregate supply function (1954, pp. 320–1).

The tatonnement in the money market, Walras contended, explains how the rate of interest and the equilibrium aggregate net quantities of cash balances supplied and demanded are determined (1954, pp. 325, 327). The rate of interest changes according to the Walrasian pricing rule. When the excess quantity demanded of cash balances is positive, the rise in the rate decreases the quantity demanded of cash balances by consumers and entrepreneurs by increasing the cost of the service

of availability that money provides, and also decreases the quantity demanded by entrepreneurs by causing a fall in profits and hence in the desired rate of production. The rise in the rate of interest also causes the net quantity of cash balances that savers want to supply to increase. If the desired supply of cash exceeds the desired demand at the current rate of interest, the opposite effects occur (1954, p. 333). The tatonnement continues until the equilibrium price of the service of availability of money is found -namely, the price that equates the net and therefore the gross quantities supplied and demanded of cash balances – whereupon the money market reaches equilibrium (1899, p. 96; 1900, pp. 297–319; 1954, pp. 315–33).

The equilibrium prices of all goods in terms of money are given by its role as the numeraire and by the workings of the entire model that determine the ratio of exchange between each good and the numeraire. In general equilibrium, the price of the service of all money held by different individuals for different purposes is the same (1954, p. 326). Moreover, because an underlying influence upon the rate of interest on money is the value productivity of physical capital, an influence exerted through variations in the volume of funds invested, the equilibrium rate on money is the same as the equilibrium rate of net income determined in the market for capital. There is therefore equality in the rate of net income from all capital goods and real and monetary circulating capital (1954, p. 323).

Walras then considered the comparative statics of the model. He changed the utility functions for the service of money and deduced that the marginal utility and value of the service of money changes in the same direction. He changed the quantity of money and deduced that the marginal utility and value of the service of money changes in the opposite direction, and that all prices change in the same direction without any alterations in relative prices (1954, p. 333). He noted that, if the utility curves for net income shift up or down, the rate of net income changes in the opposite direction. If the quantity of net income varies, the rate of net income varies in the same direction. If utility functions and the quantity of net income both vary in such a way that the marginal utilities remain unchanged, the rate of net income also remains unchanged (1954, p. 307).

An aura of unreality is imparted to Walras's 1900 edition of the *Eléments* by his abstracting from money through much of his exposition of exchange, production and capital formation, and then by introducing it in such a way that it does not change their characteristics (1954, pp. 319–24). In particular, by postulating that there is no uncertainty in his last model of money and circulating capital, without the slightest explanation of how that would be possible, he eliminated consideration of much of the behaviour associated with money in the real economy. Money does, however, influence a great deal of real economic behaviour in special ways associated with uncertainty, a fact of which Walras's extensive writings on real monetary arrangements, problems and policies reveal him to have been perfectly cognizant. Moreover, his concept of fictional perpetual annuity shares is a superfluity that further detracts from the verisimilitude of his models of capital formation and money. He should instead have retained his mature model of the money market in which he dealt with the behaviour related to some of the major financial assets in which people actually invest.

Economic Policies

Walras developed all his policy proposals during the years prior to his last theoretical efforts. He never mentioned the latter formulations in connection with real empirical matters. In particular, the written pledges sketch did not influence him to modify or innovate policy proposals, necessarily so because the functioning and hence the problems of the economy are not virtual.

Walras was greatly interested in the economic problems of his day and in socioeconomic reform, guided in his major policy proposals by his normative convictions, which were derived from his father's philosophy of society and justice. Those convictions were a mixture of conventional nineteenth-century liberalism and notions of the rightness and efficacy of state interventionism

(1896b). Like many scholars, each with different views, Walras bestowed the title of 'natural law' upon the principles of justice that he considered desirable, and so he might be called a natural-law philosopher or casuist. Nevertheless, he was not a natural-law economist. He did not believe that there is, behind observable facts, a structure of economic laws that are divinely ordained, or that are peculiarly in tune with the structure of the universe and human aspirations, or are ceaselessly at work so that violations of them can only result in chaos or frictions. Nor did he construct his economic model with the conscious intention of expressing his normative views. Sharply distinguishing normative and positive economics, he stated that he designed his theories for the purpose of understanding economic reality (Walker 1984a, 2006a) and presented his normative work explicitly as such and carefully segregated it (Walras 1896b) from the publications presenting his economic theories.

Walras's policy recommendations ranged over natural monopolies, which he believed should be nationalized; prices, which he believed should be stabilized by a monetary authority; bimetallism, which he believed had both advantages and disadvantages; the stock market, which he believed should be regulated by the state in order to improve its organization and ensure its integrity; taxes, which he believed were unjust and confiscatory and should be abolished; and land, which he believed should be purchased by the state and rented to private users, thereby providing it with revenue (1905b, pp. 272–3). Arguing that his advocacy of nationalization of land and natural monopolies was based upon a scientific understanding of the functioning of the economy, Walras called himself a 'scientific socialist'.

Contributions

Criticisms of Walras's work cannot obscure the greatness of his contributions. When he began his investigations in 1868, economics on the Continent was hardly a scientific pursuit but rather a mixture of normative prescriptions, classical theories expressed alongside protectionist doctrines, and commercial law. In England it was in the state exemplified by the work of J.S. Mill –with much that could be used as a basis for future investigations, but also without a clear view of the relationships of distribution and production, limited by a cost-of-production theory of value, and lacking a theory of supply and demand in multiple markets. The attitude of most of Walras's contemporaries was that, since economic behaviour involves preferences and the human will, it cannot be expressed in a rigid and deterministic set of algebraic relations. Walras changed all that, transforming economics and propelling it forward in a gigantic intellectual leap.

His contribution can be divided into two interrelated parts. One is that, in his mature comprehensive model, he constructed or refined or adapted to his purposes many of what became the fundamental building blocks of modern economic theory. In this effort he accomplished an enormous amount of highly creative economic analysis, brilliantly analysing the structure of economic reality to bring many of its essential features into clear relief, in eight major original contributions. First, he went far beyond the work of the other developers of the marginal utility theory by using it to analyse the disequilibrium and equilibrium behaviour of a variety of participants undertaking different economic functions in multiple markets, rather than confining the theory to the investigation of consumer demand and of exchange in an isolated market. Second, he had clear priority in constructing the theory of exchange in multiple competitive markets. In that regard, his work was greatly in advance of his predecessors' and was replete with fruitful constructions, theorems and postulates, like the reciprocal relation of supply and demand, the device of a numeraire, the individual budget equation, Walras's Law, the theorem of equivalent distributions, and the laws of change of prices. Third, he constructed a theory of the firm and of market supply in which he appears to have developed independently the modern idea of a firm's production function, derived the equation for a firm's average cost, expressed the firm's offer of output mathematically, and aggregated the firm's supply functions to obtain the market

supply in a particular industry. Fourth, he was the first to examine the question of the existence of equilibrium in a competitive multi-market system of exchange and production. Fifth, in his work on tatonnement he initiated the study of the stability of competitive general equilibrium and contributed significantly to its understanding, with his most successful theorizing on the topic relating to his mature comprehensive model. There is nothing in the literature before Walras's time or until the time that his work was used by others that is even remotely like or on the level of his reasoning regarding the process of convergence to equilibrium of a non-virtual competitive multi-market system. Sixth, he developed a theory of the entrepreneur, of profits, and of the allocation of resources that became the basis of Continental work on those topics (Pareto 1896–7, passim; Pareto 1906, passim; Barone 1896, p. 145; Schumpeter 1912/1926, p. 76; Schumpeter 1954, p. 893; Walker 1986). Seventh, Walras created a fruitful theory of capital, achieving an early formulation of the conditions for a Pareto optimum in capital markets. As in a number of his other investigations, his characteristic contribution in that regard was not to be the first to think of the problem but to be the first to offer an account of those markets' disequilibrium interrelationships and equilibrium conditions in a model of the general equilibrium of an economy. Eighth, he developed a cash-balances theory of money in his period of maturity as a theoretician which had great originality and has stimulated much valuable research (Marget 1931, 1935; and see Walker 1970, p. 696; 1996, pp. 235–55). Those eight areas of analysis were the core of neoclassical microeconomic theory and thus constituted much of the structure of knowledge that was the starting place for twentieth-century economics.

The second part of Walras's contribution was not the idea of the general equilibrium of a freely competitive economic system, which other economists had suggested; it was his implementation of that idea. Other economists had helped in fashioning the building blocks that Walras used. His achievement, however, was not only to develop them but also, through incisive analysis and constructive thought, to weave them into an account of the equilibrating processes of a complex, non-virtual, multi-market economy. Those building blocks dealt with real economic behaviour, and it is his use of building blocks with that essential quality that gives his work its richness, robustness and relevance. Walras was also the first economist to try to set up a system of equations to describe the conditions of static equilibrium of a general equilibrium model.

Walras thus accomplished by the mid-1870s far more than any other economist had done in regard to constructing a model of an economic system as a whole, and more single-handedly in that regard than any other economist in the history of the discipline.

Influence

Walras's work was not given the recognition that it merited in France during the 25 years after 1874, and his centennial, in 1934, elicited no conference on his work there. By the 1950s, however, the French attitude toward Walras had changed, as was ultimately symbolized by the creation in 1984 of the Centre Auguste et Léon Walras (but no longer symbolized in that way, for the research group has a new name; it is now a part of the organization known as 'TRIANGLE, Unité mixte de Recherche 5206 du Centre national de la Recherche scientifique') at the université Lumière Lyon 2. With the English, Walras's experience was also disappointing. His initial cordiality towards W.S. Jevons, as a fellow pioneer in mathematical economics, was dissipated by Jevons's failure to recognize Walras's contributions to the theory of exchange and to the construction of a relatively complete model of a competitive economy, and eventually Walras, quite unreasonably, came to regard Jevons as a plagiarist of his work (Walras to M. Pantaleoni, 17 August 1889, L 909). Similarly, Walras's relations with P.H. Wicksteed began well (Wicksteed to Walras, 1 December 1884, L 619) but deteriorated sharply when Wicksteed failed to give credit to those whom Walras considered to be the true originators of the theory of marginal productivity

(Walras 1965, L 1220, n. 3; 1896a, pp. 490–2). Walras justly felt neglected by Alfred Marshall, who mentioned him only thrice in the briefest of comments in the *Principles* (Marshall 1890, 1920) and wrote not a word about Walras's development of general equilibrium theory. Walras also came to dislike Edgeworth for criticizing his theories of tatonnement, capital goods and the entrepreneur (Walras to Gide, 3 November 1889, L 933, and 11 April 1891, L 1000; Walras to Pantaleoni, 5 January 1890, L 953). In general, Walras believed, the English had closed their minds to his theories and had become spiteful in their treatment of them (see Walker 1970, pp. 699–70).

The extremity of the language with which Walras characterized the English was unjustified, because, although he had reason for disappointment with their neglect of his general equilibrium theory, Jevons (1879, preface) and Edgeworth (1889) had recognized valuable elements in his work, and he was the only living economist included in the first edition of Palgrave's *Dictionary of Political Economy* (Sanger 1899). The fact is that Walras grew hypersensitive about the motives of his critics, the failure of the majority of economists to recognize the value and priority of his contributions, and the possibility of plagiarism of his ideas during the 1880s and 1890s. There had been two periods in his life, he complained, 'one during which I was a madman, and one during which everyone made my discoveries before me' (Walras, undated, in Jaffé 1983, p. 203, n.54).

This account of Walras's disappointments should be balanced by a realization that his scientific labours had afforded him 'up to a certain point, pleasures and joys like those that religion provides to the faithful' (Walras to Marie de Sainte Beuve, 15 December 1899, L 1432), and a recognition of the professional satisfactions that he increasingly experienced in the last two decades of his life. Maffeo Pantaleoni (1889), Enrico Barone (1895, in Jaffé 1983, p. 186; 1896), and Vilfredo Pareto (Pareto to Walras, 15 October 1892, L 1077) contributed greatly toward giving Walras's work a secure place in Continental economics and thus ultimately in economics everywhere. In 1895 Pareto's appointment as Walras's successor to the chair of economics at Lausanne assured Walras that his doctrines, expressed in his treatment of a non-virtual competitive economy, would be perpetuated and developed, and the accessible literary presentations of Walras's ideas in Pareto's books (1896–7; 1906) began their widespread dissemination. Pareto borrowed most of the ideas of Walras that have been mentioned in this article, using them as the basis for his contributions to the theories of non-virtual general equilibrium, the monopolistic entrepreneur, and welfare economics. Wilhelm Lexis, Ladislaus von Bortkiewicz and Eugen von Bohm-Bawerk gave Walras's models serious attention. Knut Wicksell based his theory of price determination squarely upon Walras's work (Wicksell to Walras, 6 November 1893, L 1168), and Karl Gustav Cassel was inspired in the construction of his models (1903, 1918) by Walras's equation system and idea of virtuality expressed in the written pledges sketch. Walras was given recognition in the United States: in 1892 he was made an honorary member of the American Economic Association, Irving Fisher praised his work (Fisher 1892, p. 45; 1896), H.L. Moore became his avowed disciple and explicator (Moore to Walras, 19 May 1909, enclosure to L 1747; Moore 1929), and Henry Schultz taught Walras's economics and, like Moore, undertook theoretical and econometric studies of general equilibrium relationships (for example, Schultz 1929, 1932, 1933). His mature comprehensive model was the starting point for the work of Joseph Schumpeter, who, throughout his entire career, devoted himself to the study of disequilibrium and general equilibration of non-virtual economic phenomena.

The manifestations of acceptance led Walras to believe he would ultimately triumph, and that enabled him to achieve a mental calmness (Walras to Marie de Sainte Beuve, 15 December 1899, L 1432; Walras to A. Aupetit, 28 May 1901, L 1485). 'Be assured of my serenity', he wrote to old friends in 1904, 'I have not the least doubt about the future of my method and even of my doctrine; but I know that success of this sort does not become clearly apparent until after the death of the author'

(Walras to G. and L. Renard, 4 June 1904, L 1574). Walras's prediction of great success was accurate. An indication of what the future would hold for his theories was given by the celebration of his jubilee in 1909 by the University of Lausanne, in the course of which he was honoured as the first economist to establish the conditions of general equilibrium, thus founding the School of Lausanne (Walras 1965, L 1696, n. 5). His achievements were praised in a statement signed by 15 leading French scholars, including Charles Gide, Charles Rist, Georges Renard, Alfred Bonnet, Albert Aupetit and François Simiand (Walras 1965, enclosure to L 1747), and in communications from many others (Pareto to the Dean of the Faculty of Law of the University of Lausanne, 6 June 1909, L 1755; Schumpeter to Walras, 7 June 1909, L 1756).

Walras's contributions inspired and provided a substantial beginning for the branches of general equilibrium theory and applications as they have developed since the nineteenth century. Indeed, the filiations of his ideas have become so numerous and dense as to be an integral and central part of the mainstream of modern economics. His achievement of developing particular theories and binding them together in a model of an entire economic system has given his work an influence on the verbal, mathematical and econometric study of the interrelationships of all parts of economic systems that has been durable and immense. For sheer genius and intuitive power in penetrating the veil of the chaos of immediately perceived experience and divining the underlying structure of fundamental economic relationships and their extensive interdependencies and consequences, Walras has been surpassed by no one.

See Also

- ▶ General Equilibrium
- ▶ Tâtonnement and Recontracting

Selected Works

1858. *Francis Sauveur.* Paris: E. Dentu.

1860. *L'Economie politique et la justice.* Paris: Guillaumin.

1868. *Recherche de l'idéal social: Leçons publiques faites à Paris.* Paris: Guillaumin; and in Walras (1896a).

1869–70. 2[e] tentative, 1869–1870. In Walras (1965), vol. 1, 218–219.

1871. *Discours d'installation. Séance académique du 20 Octobre.* Lausanne: Académie de Lausanne.

1877a. *Théorie mathématique de la richesse sociale: quatre mémoires.* Paris: Guillaumin.

1874, 1877b. *Eléments d'économie politique pure; ou théorie de la richesse sociale.* 2 parts. Lausanne/Paris/Bâsle: L. Corbaz/Guillaumin/H. Georg.

1880a, 1880b. La bourse, la spéculation et l'agiotage. *Bibliothéque Universelle et Revue Suisse* 5: March 1880a, 452–476; 6: April 1880b, 66–94.

1886. *Théorie de la monnaie.* Lausanne: Corbaz.

1889. *Eléments d'économie politique pure,* 2nd edn. Lausanne/Paris/Leipzig: F. Rouge/Guillaumin/Duncker & Humblot.

1893, 1904, 1909. Notice autobiographique. In Walras (1965), vol. 1, 1–15.

1895. Enclosure by Léon Walras to Vilfredo Pareto. In Walras (1965), vol. 2, 628–632.

1896a. *Eléments d'économie politique pure,* 3rd edn. Lausanne/Paris/Leipzig: F. Rouge/F. Pichon/Duncker & Humblot.

1896b. *Etudes d'économie sociale* (*Théorie de la réparation de la richesse sociale*). Lausanne/Paris: F. Rouge/F. Pichon.

1898. *Etudes d'économie politique appliquée (Théorie de la production de la richesse sociale)*. Lausanne/Paris: F. Rouge/F. Pichon.

1899. Equations de la circulation. *Bulletin de la Société Vaudoise des Sciences Naturelles,* June, 85–102 and note, 103.

1900. *Eléments d'économie politique pure,* 4th edn. Lausanne/Paris: F. Rouge/F. Pichon.

1905a. Cournot et l'économique mathématique. *Gazette de Lausanne,* 13 July; and, except for a few deletions, in *Revue d'Economie Politique,* January–February, 1962.

1905b. Draft of a recommendation for a Nobel Peace Prize. In Walras (1965), vol. 3, 270–274.

1926. *Eléments d'économie politique pure.* Definitive edn. Paris/Lausanne: R. Picon & R. Durand-Auzias/F. Rouge.

1954. *Elements of pure economics.* Trans. and annotated by William Jaffé. Homewood/London: Richard D. Irwin/Allen & Unwin.

1965. *Correspondence of Léon Walras and related papers,* 3 vols, ed. W. Jaffé. Amsterdam: North-Holland.

Bibliography

Barone, E. 1895. *On a recent book by Wicksteed. Manuscript translated into English by Walker in Jaffé* (1983), 182–186.

Barone, E. 1896. Studi sulla distribuzione. *Giornale degli Economisti* 12(February): 107–155, 235–252.

Bertrand, J. 1883. Théorie des richesses. *Journal des Savants* 67 (September): 499–508.

Cassel, G. 1903. *Nature and necessity of interest*. London: Macmillan.

Cassel, G. 1918. *Theory of social economy.* New York: Harcourt, 1932.

Cournot, A.A. 1838. *Recherches sur les principes mathématiques de la théorie des richesses*. Paris: Hachette.

Edgeworth, F.Y. 1889. The mathematical theory of political economy. [Review of] *Eléments d'économie politique pure* by Léon Walras. *Nature* 40: 434–436.

Fisher, I. 1892. Translator's note to Léon Walras, Geometrical theory of the determination of prices. *Annals of the American Academy of Political and Social Science* 3 (July): 45–47.

Fisher, I. 1896. [Review of] *Eléments d'économie politique pure. Troisième edition. Par Léon Walras* . . .. [and of] *An essay on the co-ordination of the laws of distribution. By Philip H Wicksteed. . .. Yale Review* 5 (1): 222–223.

Hicks, J.R. 1934. Léon Walras. *Econometrica* 2: 338–348.

Isnard, A.N. 1781. *Traité des richesses, contenant l'analyse de l'usage des richesses en général et de leurs valeurs*. London/Lausanne: Grasset.

Jaffé, W. 1964. New light on an old quarrel: Barone's unpublished review of Wicksteed's essay on the coordination of the laws of distribution and related documents. *Cahiers Vilfredo Pareto* 3: 61–102; where necessary translated into English by Walker in Jaffé (1983).

Jaffé, W. 1969. A.N. Isnard, progenitor of the Walrasian general equilibrium model. *History of Political Economy* 1: 19–43; and in Jaffé (1983).

Jaffé, W. 1971. Reflections on the importance of Léon Walras. In the P. Hennipman festschrift, *Schaarste en Welvaart,* ed. A. Heertje et al. Amsterdam: Stenfert Kroese; and in Jaffé (1983).

Jaffé, W. 1981. Another look at Léon Walras's theory of *tâtonnement. History of Political Economy* 13: 313–336; and in Jaffé (1983).

Jaffé, W. 1983. *William Jaffé's essays on Walras,* ed. D.A. Walker. Cambridge/New York: Cambridge University Press.

Jevons, W.S. 1879. *The theory of political economy.* 2nd ed. London: Macmillan.

Klotz, G. 1994. Achylle Nicolas Isnard, précurseur de Léon Walras? *Economies et Sociétés, Série Œconomia, Histoire de la Pensée économique* 28 (10–11): 29–52.

Lexis, W. 1881. Zur mathematisch-ökonomischen Literatur. *Jahrbücher für Nationalökonomie und Statistik* 3: n.s.427–434.

Marget, A.W. 1931. Léon Walras and the 'cash-balance approach' to the problem of the value of money. *Journal of Political Economy* 39: 569–600.

Marget, A.W. 1935. The monetary aspects of the Walrasian system. *Journal of Political Economy* 43: 145–186.

Marshall, A. 1890. *Principles of economics*, 8th ed. London: Macmillan, 1920.

Moore, H.L. 1929. *Synthetic economics*. New York: Macmillan.

Pantaleoni, M. 1889. *Principii de economia pura*. Florence: Barbèra.

Pareto, V. 1896–7. *Cours d'économie politique,* 2 vols. Lausanne: F. Rouge. Ed. G.-H. Bousquet and G. Busino, Geneva: Librairie Droz, 1964.

Pareto, V. 1906. *Manuale di economia politica.* Milan: Società Editrice Libraria. Trans. A. Bonnet from the Italian as *Manuel d'économie politique.* Paris: V. Giard and E. Brière, 1909.

Poinsot, L. 1803. *Eléments de statique*, 8th ed. Paris: Bachelier, 1842.

Sanger, Cs.P. 1899. Walras, Marie Esprit Léon. In *Dictionary of political economy*, ed. R.H. Inglis Palgrave, vol. 3. London: Macmillan.

Say, J.-B. 1836. *Cours complet d'économie politique pratique, suivi des Melanges, Correspondance et Catéchisme d'économie politique*. 3rd ed. Bruxelles: H. Dumont.

Schultz, H. 1929. Marginal productivity and the general pricing process. *Journal of Political Economy* 37: 505–551.

Schultz, H. 1932. Marginal productivity and the Lausanne School. *Economica* 12: 285–296.

Schultz, H. 1933. Interrelations of demand. *Journal of Political Economy* 41: 468–512.

Schumpeter, J.A. 1912, 1926. *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle.* Trans. R. Opie. New York: Oxford University Press, 1961.

Schumpeter, J.H. 1954. *History of economic analysis*. New York: Oxford University Press.

Van Daal, J., R.E.D. Henderiks, and A.C.F. Vorst. 1985. On Walras' model of general economic equilibrium. *Zeitschrift für Nationalökonomie* 45 (3): 219–244.

Walker, D.A. 1970. Léon Walras in the light of his correspondence and related papers. *Journal of Political Economy* 78(4), Part I: 685–701.

Walker, D.A. 1984a. Is Walras's theory of general equilibrium a normative scheme? *History of Political Economy* 16: 445–469.

Walker, D.A. 1984b. Walras and his critics on the maximum utility of new capital goods. *History of Political Economy* 16: 529–554.

Walker, D.A. 1986. Walras's theory of the entrepreneur. *De Economist* 134 (1): 1–24.
Walker, D.A. 1996. *Walras's market models*. Cambridge/New York/Melbourne: Cambridge University Press.
Walker, D.A. 2000. La relation entre la Bourse au XIX^{e} siècle et le modèle de Léon Walras d'un marché organisé. In *Les traditions économiques françaises 1848–1939*, ed. P. Dockès et al. Paris: Éditions du CNRS.
Walker, D.A. 2001. A factual account of the functioning of the nineteenth-century Paris Bourse. *European Journal of the History of Economic Thought* 8: 186–207.
Walker, D.A. 2006a. *Walrasian economics*. Cambridge/New York: Cambridge University Press.
Walker, D.A. 2006b. *Virtualité et non virtualité dans les modèles de l'équilibre général*. Colloque Walras, Association Internationale Walras, University of Lausanne, 22 September.
Walras, A.A. 1831. *De la nature de la richesse et de l'origine de la valeur*. Paris: Johanneau.
Weintraub, E.R. 1983. On the existence of a competitive equilibrium: 1930–1954. *Journal of Economic Literature* 21: 1–39.

Walras's Law

Don Patinkin

Abstract

Walras's Law is an expression of the interdependence among the excess-demand equations of a general-equilibrium system that stems from the budget constraint. It is so called because Walras made use of this interdependence from the first edition of his *Eléments d'économie politique pure* through the fourth, which for all practical purposes is identical with the definitive edition. Walras presents the argument for an exchange economy and extends his analysis to deal first with a simple production economy and then with one in which capital formation also takes place. Here, Walras's Law is derived in a more general way than usual.

Keywords

Barter; Dual decision hypothesis; Excess demand and supply; General equilibrium; Hicks, J. R.; Keynes, J. M.; Kinked demand curve; Kinked supply curve; Lange, O. R.; Liquidity preference; Loanable funds; Modigliani, F.; Patinkin, D.; Perfect competition; Price rigidity; Real-balance effect; Ricardo, D.; Say, J.-B.; Say's Law; Tâtonnement; Wage rigidity; Walras's law

JEL Classifications

D5

Walras's Law (so named by Lange 1942) is an expression of the interdependence among the excess-demand equations of a general-equilibrium system that stems from the budget constraint. Its name reflects the fact that Walras, the father of general-equilibrium economics, himself made use of this interdependence from the first edition of his *Eléments d'économie politique pure* (1874, §122) through the fourth (1900, §116), which edition is for all practical purposes identical with the definitive one (1926). I have cited §116 of this edition because it is the one cited by Lange (1942, p. 51, n. 2), though in a broader context than Walras's own discussion there (see below). In this section, Walras presents the argument for an exchange economy. In accordance with his usual expository technique (cf. his treatment of the *tâtonnement*), he repeats the argument as he successively extends his analysis to deal first with a simple production economy and then with one in which capital formation also takes place (ibid., §§ 206 and 250, respectively).

For reasons that will become clear later, I shall derive Walras's Law in a more general – and more cumbersome – way than it usually has been. Basically, however, the derivation follows that of Arrow and Hahn (1971, pp. 17– 21), with an admixture of Lange (1942) and Patinkin (1956, chs I– III and Mathematical Appendix 3:a).

Let x_i^h be the decision of household h with respect to good $i(i = 1,\ldots, n)$, where 'goods' also include services and financial assets (securities and money). If $x_i^h \geq 0$, it is a good purchased by the household; if $x_i^h < 0$, it is a good (mainly, labour or some other factor-service)

sold. Similarly, let y_i^f be the decision of firm f with respect to good i; if $y_i^f \geq 0$, it is a good produced and sold by the firm (i.e., a product-output); if $y_i^f < 0$, it is a factor-input.

Assume that firm f has certain initial conditions (say, quantities of fixed factors of production) represented by the vector $\mathbf{k}^f$ and operates in accordance with a certain production function. Following Patinkin (1956), let us conduct the conceptual individual-experiment of confronting the firm with the vector of variables $\mathbf{v}$ (the nature of which will be discussed below) while keeping $\mathbf{k}^f$ constant and asking it to designate (subject to its production function) the amounts that it will sell or buy of the various goods and services. By repeating this conceptual experiment with different values of the respective elements of $\mathbf{v}$, we obtain the behaviour functions of firm f,

$$y_i^f = y_i^f\left(\mathbf{v}; \mathbf{k}^f\right) \ (i = 1, \ldots, n). \tag{1}$$

For $y_i \geq 0$, this is a supply function; for $y_i < 0$, it is a demand function for the services of factors of production. Profits (positive or negative) of firm f are then

$$R^f = \sum_i p_i y_i^f\left(\mathbf{v}; \mathbf{k}^f\right), \tag{2}$$

Let d^{hf} represent the proportion of the profits of firm f received by the household h. Its total profits received are then $\Sigma_f d^{hf} R^f$ and its budget constraint is accordingly

$$\sum_i p_i x_i^h = \sum_f d^{hf} R^f, \tag{3}$$

which assumes that households correctly estimate the profits of firms (cf. Buiter 1980, p. 7; I shall return to this point below). As with the firm, let us, *mutatis mutandis*, conduct individual-experiments with household h (with its given tastes), subject to its budget constraint (3) by varying the elements of $\mathbf{v}$, while keeping its initial endowment (represented by the vector $\mathbf{e}^h$ constant. This yields the behaviour functions

$$x_i^h = x_i^h\left(\mathbf{v}; \mathbf{e}^h\right) \ (i = 1, \ldots, n). \tag{4}$$

For $x_i \geq 0$, this is a demand function for goods; for $x_i < 0$, it is a supply function (e.g., of factor-services).

Substituting from (2) and (4) into (3) then yields

$$\sum_i p_i x_i^h\left(\mathbf{v}; \mathbf{e}^h\right) = \sum_f d^{hf} \sum_i p_i y_i^f\left(\mathbf{v}; \mathbf{k}^f\right), \tag{5}$$

which holds identically for all $\mathbf{v}$, $\mathbf{e}^h$, $\mathbf{k}^f$ and p_i. Summing up over all households then yields

$$\sum_h \sum_i p_i x_i^h\left(\mathbf{v}; \mathbf{e}^h\right) = \sum_h \times \sum_f d^{hf} \sum_i p_i y_i^f\left(\mathbf{v}; \mathbf{k}^f\right), \tag{6}$$

which we rewrite as

$$\sum_i p_i \sum_h x_i{}^h\left(\mathbf{v}; \mathbf{e}^h\right) = \sum_i p_i \sum_f \left(\sum_h d^{hf}\right) y_i^f\left(\mathbf{v}; \mathbf{k}^f\right). \tag{7}$$

On the assumption that firm f distributes all its profits,

$$\sum_h d^{hf} = 1 \text{ for all } f, \tag{8}$$

so that (7) reduces to

$$\sum_i p_i [X_i(\mathbf{v}; \mathbf{E}) - Y_i(\mathbf{v}; \mathbf{K})] = 0 \tag{9}$$

identically in all $\mathbf{v}$, $\mathbf{E}$, $\mathbf{K}$ and p_i, where

$$X_i(\mathbf{v}; \mathbf{E}) = \sum_h x_i^h\left(\mathbf{v}; \mathbf{e}^h\right) \text{ and } Y_i(\mathbf{v}; \mathbf{K}) = \sum_f y_i^f\left(\mathbf{v}; \mathbf{k}^f\right) \tag{10}$$

represent the aggregate demand and supply functions, respectively, for good i; $\mathbf{E}$ is a vector

containing all the $\mathbf{e}^h$; and $\mathbf{K}$ a vector containing all the $\mathbf{k}^f$. If $X_i(\mathbf{v};\ \mathbf{E}) - Y_i(\mathbf{v};\ \mathbf{K}) > 0$, an excess demand is said to exist in the market; if $X_i(\mathbf{v};\ \mathbf{E}) - Y_i(\mathbf{v};\ \mathbf{K}) < 0$, an excess supply; and if $X_i(\mathbf{v};\ \mathbf{E}) = Y_i(\mathbf{v};\ \mathbf{K})$, equilibrium.

Equation (9) is a general statement of Walras's Law. Its most frequent application in the literature has been (as in Walras's *Eléments* itself) to the general-equilibrium analysis of a system of perfect competition, in which the behaviour functions of firms are derived from the assumption that they maximize profits subject to their production function; and those of households are derived from the assumption that they maximize utility subject to their budget constraint. In this context, the vector $\mathbf{v}$ is the price vector $(p_1, \ \ldots\ , p_{n-1})$, with the nth good being money and serving as numéraire (i.e., $p_n = 1$), so that there are only $n - 1$ prices to be determined. Ignoring for simplicity vectors $\mathbf{E}$ and $\mathbf{K}$, which remain constant in the conceptual market-experiment, Eq. (9) then becomes

$$\sum_{i=1}^{n} p_i \left[X_i(p_1, \ldots, p_{n-1}) - Y_i(p_1, \ldots, p_{n-1})\right] = 0 \ \textit{identically in the}\ p_i. \tag{11}$$

(Though it does not bear on the present subject, I should note that under the foregoing assumptions, and in the absence of money illusion, each of the demand and supply functions is homogeneous of degree zero in $p_1\ ,\ \ldots\ ,\ p_{n-1}$ and in whatever nominal financial assets are included in $\mathbf{E}$ and $\mathbf{K}$ (e.g., initial money holdings).) Thus Walras's Law states that no matter what the p_i, the aggregate value of excess demands in the system equals the aggregate value of excess supplies. This is the statement implicit in Lange's presentation (1942, p. 50).

Walras himself, however, sufficed with a particular and narrower application of this statement, and was followed in this by, *inter alia*, Hicks (1939, chs. IV: 3 and XII: 4– 5), Modigliani (1944, pp. 215– 16) and Patinkin (1956, ch. III: 1– 3). Assume that it has been shown that a certain price vector $\left(p_1^0, \ldots, p_{n-1}^0\right)$ equilibrates all markets but the jth. Since (11) holds identically in the p_i, it must hold for this price vector too. Hence substituting the $n - 1$ equilibrium conditions into (11) reduces it to

$$p_j^0 \left[X_j\left(p_1^0, \ldots, p_{n-1}^0\right) - Y_j\left(p_1^0, \ldots, p_{n-1}^0\right)\right] = 0. \tag{12}$$

Thus if $p_j^0 > 0$, the price vector $\left(p_1^0, \ldots, p_{n-1}^0\right)$ must also equilibrate the jth market, which means that only $n - 1$ of the equilibrium equations are independent. In this way Walras (and those who followed him) established the equality between the number of independent equations and the number of price-variables to be determined. (Though such an equality is not a sufficient condition for the existence of a unique solution with positive prices, it is a necessary – though not sufficient – condition for the peace-of-mind of those of us who do not aspire to the rigour of mathematical economists.)

It should however be noted that at the end of §126 of Walras's *Eléments* (1926), there is a hint of Lange's broader statement of the Law: for there Walras states that if at a certain set of prices 'the total demand for some commodities is greater (or smaller) than their offer, then the offer of some of the other commodities must be greater (or smaller) than the demand for them'; what is missing here is the quantitative statement that the respective aggregate values of these excesses must be equal.

Since the contrary impression might be gained from some of the earlier literature (cf., e.g., Modigliani 1944, pp. 215–16), it should be emphasized that no substantive difference can arise from the choice of the equation to be 'dropped' or 'eliminated' from a general-equilibrium system by virtue of Walras's Law. For identity (11) can be rewritten as

$$X_j(p_1, \ldots, p_{n-1}) - Y_j(p_1, \ldots, p_{n-1}) = -\frac{1}{p_j} \sum_{\substack{i=1 \\ i \neq j}}^{n} p_i [X_i(p_1, \ldots, p_{n-1}) - Y_i(p_1, \ldots, p_{n-1})] \ \textit{identically in the}\ p_i.$$

Thus the properties of the 'eliminated' equation are completely reflected in the remaining ones.

Correspondingly, no matter what equation is 'eliminated', the solution for the equilibrium set of prices obtained from the remaining equations must be the same. (From this it is also clear that the heated 'loanable-funds versus liquidity-preference' debate that occupied the profession for many years after the appearance of the *General Theory*, was largely misguided; see Hicks 1939, pp. 157–62; see also Patinkin 1956, ch. XV: 3, and 1958, pp. 300–302, 316–17.)

In his influential article, Lange (1942, pp. 52– 53) also distinguished between Walras's Law and what he called Say's Law, and I digress briefly to discuss this. As before, let the first $n-1$ goods represent commodities and the nth good money. Then Say's Law according to Lange is

$$\sum_{i=1}^{n-1} p_i X_i(p_1, \ldots, p_{n-1}) = \sum_{i=1}^{n-1} p_i Y_i(p_1, \ldots, p_{n-1}) \; \textit{identically in the} \; p_i. \tag{13}$$

That is, the aggregate value of commodities supplied at any price vector $(p_1, \ldots, p_{n-1})$ must equal the aggregate value demanded: supply always creates its own demand.

On both theoretical and doctrinal grounds, however, I must reject Lange's treatment of Say's Law. First of all, Lange himself demonstrates (ibid., p. 62) that identity (13) implies that money prices are indeterminate. In particular, subtracting (13) from (11) yields

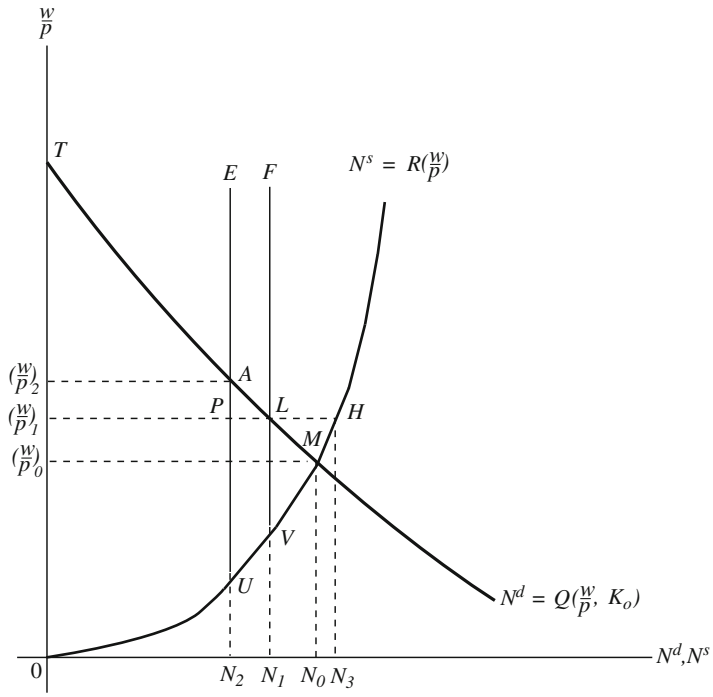
$$X_n(p_1, \ldots, p_{n-1}) = Y_n(p_1, \ldots, p_{n-1}) \text{ for all } p_i. \tag{14}$$

That is, no matter what the price vector, the excess-demand equation for money must be satisfied, which in turn implies that money prices cannot be determined by market forces. But it is not very meaningful to speak of a money economy whose money prices are indeterminate even for fixed initial conditions as represented by the vectors **E** and **K**. So if we rule out this possibility, we can say that Say's Law in Lange's sense implies the existence of a barter economy. Conversely, in a barter economy (i.e., one in which there exist only the $n-1$ commodities) Say's Law is simply a statement of Walras's Law. Thus from the above viewpoint, a necessary and sufficient condition for the existence of Say's Law in Lange's sense is that the economy in question be a barter economy: it has no place in a money economy (Patinkin 1956, ch. VIII: 7).

Insofar as the doctrinal aspect is concerned, identity ($_{13}$) cannot be accepted as a representation of Say's actual contention. For Say's concern was (in today's terminology) not the short-run viewpoint implicit in this identity, but the viewpoint which denied that in the long run inadequacy of demand would set a limit to the expansion of output. In brief, and again in today's terminology, Say's concern was to deny the possibility of secular stagnation, not that of cyclical depression and unemployment. Thus, writing in the first quarter of the 19th century, Say (1821a, p. 137) adduces evidence in support of his thesis from the fact 'that there should now be bought and sold in France five or six times as many commodities, as in the miserable reign of Charles VI – four centuries earlier. Again, in his *Letters to Malthus* (1821b, pp. 4– 5) Say argues that the enactment of the Elizabethan Poor Laws (codified at the end of the 16th century) proves that '*there was* no employ in a country which since then has been able to furnish enough for a double and triple number of labourers' (italics in original). Similarly, Ricardo, the leading contemporary advocate of Say's *loi des débouchés*, discusses this law in chapter 21 of his *Principles* (1821), entitled 'Effects of Accumulation of Profits and Interest'; on the other hand, he clearly recognizes the short-run 'distress' that can be generated by 'Sudden Changes in the Channels of Trade' (title of ch. 19 of his *Principles*. For further discussion see Patinkin 1956, Supplementary Note L.)

Let me return now to the general statement of Walras's Law presented in Eq. (9) above. This statement holds for any vector **v** and not only for that appropriate to perfect competition. *In particular, Walras's Law holds also for the case in which households and/or firms are subject to quantity* constraints. In order to bring this out, consider the analysis of a disequilibrium economy presented in chapter XIII:2 of Patinkin (1956) and illustrated

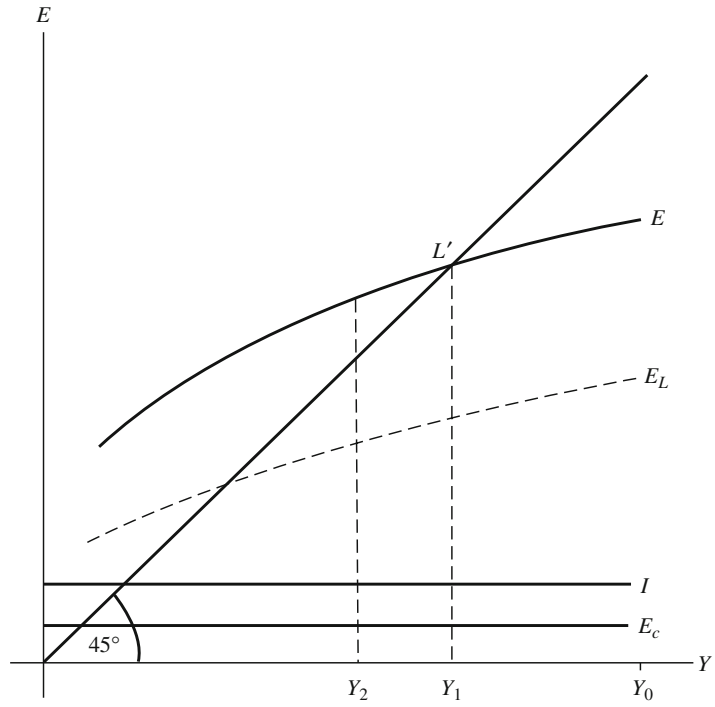
Walras's Law, Fig. 1



by Fig. 1. In this figure, w is the money wage-rate, p the price level, N the quantity of labour, $N^d = Q(w/p, K_0)$ the firms' demand curve for labour as derived from profit-maximization as of a given stock of physical capital K_0; and $N^s = R(w/p)$ is the supply curve of labour as derived from utility maximization subject to the budget constraint (these perfect-competition curves are what Clower (1965, p. 119) subsequently denoted as 'notional curves'). Assume that because of the firms' awareness that at the real wage rate $(w/p)_1$ they face a quantity constraint and will not be able to sell all of the output corresponding to their profit-maximizing input of labour N_1, they demand only N_2 units of labour, represented by point P in Fig. 1. This constraint also operates on workers, who can sell only the foregoing quantity of labour instead of their optimal one N_3, represented by point H. In brief, at point P, both firms and workers are off their notional curves. In order to depict this situation, the notional curves must accordingly be replaced by quantity-constrained ones; namely, the kinked demand curve TAN_2 and kinked supply curve OUE. Note that for levels of employment before they become kinked, the curves coincide respectively with the notional ones (but see Patinkin 1956, ch. XIII: 2, n. 9, for a basic analytical problem that arises with respect to the kinked demand curve TAN_2).

The obverse side of these constraints in the labour market are corresponding constraints in the commodity market. In particular, as Clower (1965, pp. 118–21) has emphasized, the demands of workers in this market are determined by their constrained incomes. Clower also emphasizes that it is this quantity constraint which rationalizes the consumption function of Keynes's *General Theory*, in which income appears as an independent variable. For in the absence of such a constraint, the individual's income is also a dependent variable, determined by the optimum quantity of labour he decides to sell at the given real wage rate in accordance with the labour-supply function $N^s = R(w/p)$ in Fig. 1; and he makes this decision simultaneously with the one with respect to the optimum quantities of

Walras's Law, Fig. 2



commodities to buy. If, however, his income is determined by a quantity constraint which prevents him from selling his optimum quantity of labour, the individual can decide on his demands for commodities only after his income is first determined. This is the so-called 'dual decision hypothesis' (Clower, ibid.). To this I would add (and its significance will become clear below) that the quantity constraint also rationalizes the form of Keynes's liquidity preference function, for this too depends on income (*General Theory*, p. 199). Furthermore, if the behaviour functions in the markets for labour, commodities and money balances are thus quantity-constrained, so too (by the budget constraint) will be that for bonds – the fourth market implicitly (and frequently explicitly) present in the Keynesian system. (The theory of the determination of equilibrium under quantity constraints – in brief, disequilibrium theory – has been the subject of a growing literature, most of it highly technical; for critical surveys of this literature, see Grandmont 1977; Drazen 1980; Fitoussi 1983; and Gale 1983, ch. 1.)

In the *General Theory* (ch. 2), Keynes accepted the 'first classical postulate' that the real wage is equal to the marginal product of labour, but rejected the second one, that it always also measures the marginal disutility of labour. In terms of Fig. 1 this means that while firms are always on their demand curve $N^d = Q(w/p, \mathrm{K}_0)$, workers are not always on their supply curve $N^s = R(w/p)$. Thus, for example, at the level of employment N_2, the labour market will be at point A on the labour demand curve, corresponding to the real wage rate $(w/p)_2$; but the marginal utility of the quantity of commodities that workers then buy with their real-income $(w/p)_2$. N_2 is greater than the marginal disutility of that level of employment. And Keynes emphasizes that only in a situation of full-employment equilibrium – represented by intersection point M in Fig. 1 – will both classical postulates be satisfied.

Consider now the commodity market as depicted in the usual Keynesian-cross diagram (Fig. 2). The 45° line represents the amounts of commodities which firms produce and supply as they move along their labour-demand curve from

point T to M. Thus Y represents the output (in real terms) of N_0 units of labour. Note too that the negative slope of the labour-demand curve implies that the real wage declines as we move rightwards along the 45° line.

Curve E represents the aggregate demand curve, which is the vertical sum of the consumption function of workers (E_L) and capitalists (E_c), respectively, and of the investment function (I). For simplicity, these last two are assumed to be constant. The fact that curve E does not coincide with the 45° line reflects Keynes's assumption that in a monetary economy, Say's Law (in Keynes's sense, which is the macroeconomic counterpart of Lange's subsequent formulation) does not hold (ibid., pp. 25–6).

Consider now the consumption function of workers. The income which they have at their disposal is their constrained income as determined by the labour-demand curve in Fig. 1. Thus assume that Y_1 and Y_2 in Fig. 2 are the outputs corresponding to the levels of employment N_1 and N_2, respectively. The corresponding incomes of workers at these levels are $(w/p)_1 \cdot N_1$ and $(w/p)_2 \cdot N_2$. On the assumption that the elasticity of demand for labour is greater than unity, the higher the level of employment the greater the income of workers and hence their consumption expenditures. From Fig. 2 we see that at income Y_2 there is an excess demand for commodities. This causes firms to expand their output to, say, Y_1, and hence their labour-input to, say, N_1, thus causing the constrained labour-supply curve to shift to the right to the kinked curve *OUVLF*. By construction, Y_1 is the equilibrium level of output.

What must now be emphasized is that Walras's Law holds in this situation too – *provided we relate this Law to excess-demand* functions *of the same type*. Thus if within our four-good Keynesian model we *consistently* consider notional behaviour functions, the excess supply of labour *LH* in Fig. 1 corresponds to an excess demand for commodities which is generated by workers' planned consumption expenditures at the real wage rate $(w/p)_1$ and level of employment N_3 as compared with firms' planned output at that wage rate and lower level of employment N_2; and there will generally also exist a net excess planned demand for bonds and money. Alternatively, if we *consistently* consider constrained functions, then constrained equilibria exist in the labour market (point L), the commodity market (point L'), the bond market and the money market. Similarly, the broader form of Walras's Law states that a constrained (say) excess supply in the commodity market corresponds to a constrained net excess demand in the bond and money markets, while the labour market is in constrained equilibrium. In brief, a sufficient condition for the validity of Walras's Law is that the individual's demand and supply functions on which it is ultimately based are all derived from the same budget constraint, whether quantity-constrained or not. (This is the implicit assumption of Patinkin's (1956, p. 229; 1958, pp. 314– 16) application of Walras's Law to a disequilibrium economy with unemployment, and the same is true for Grossman (1971) and Barro and Grossman (1971, 1976, p. 58).)

I must admit that the validity of Walras's Law in this Keynesian model depends on our regarding the kinked curve *OUVLF* as a labour-supply curve, and that this is not completely consistent with the usual meaning of a supply curve or function. For such a function usually describes the behaviour of an agent under constraints which leave him some degree of freedom to choose an optimum, whereas no such freedom exists in the vertical part of *OUVLF*. However, I prefer this inconsistency to what I would consider to be the logical – and hence more serious – inconsistency that lies at the base of the rejection of Walras's Law, and which consists of lumping together behaviour functions derived from different budget constraints.

It is thus clear that the foregoing constrained equilibrium in the labour market is not an equilibrium in the literal sense of representing a balance of opposing market forces, but simply the reflection of the passive adjustment by workers of the amount of labour they supply to the amount demanded by firms (cf. Patinkin 1958, pp. 314–15). From this viewpoint, the constrained equilibrium in the labour market always exists and simply expresses the fact that, by definition, every *ex post* purchase is also an *ex post* sale. In contrast, as we have seen in the discussion of Fig. 2 above, the corresponding constrained equilibrium in the commodity market

is a true one; for, in accordance with the usual Keynesian analysis, were the level of Y to deviate from Y_1, automatic market forces of excess demand or supply would be generated that would return it to Y_1. And a similar statement holds, *mutatis mutandis*, for the constrained equilibria in the bond and money markets.

Note, however, that in the commodity market too there is an *ex post* element. This element is a basic, if inadequately recognized, aspect of the household behaviour implied by Clower's 'dual decision hypothesis': namely, that households' constrained decisions on the amount of money to spend on commodities is based on their *ex post* knowledge of the amount of money received from the constrained sale of their factor services. And to this I again add that a similar statement holds for their constrained decisions with reference to the amounts of bonds and money balances, respectively, that they will want to hold. (Note that an analysis in terms of constrained decisions can also be applied to the case in which households do not correctly estimate firms' profits in Eq. (3) above, and are thus forced to base their effective (say) consumption decisions on the *ex post* knowledge of these profits.)

In this treatment of an economy with constrained functions, Clower (1965, pp. 122–3) has claimed that under these conditions Walras's Law does not hold. This is not true for the Law as hitherto discussed. What Clower seems to have in mind, however, is that though the excess supply of labour LH in Fig. 1 is notional, it nevertheless exerts pressure on workers to reduce their money wages; in contrast, the notional excess demand for commodities corresponding to LH (see above) cannot – because of their constrained incomes – lead households to exert expansionary pressures on the commodity market. Thus there exists no *effective* excess demand for commodities to match the *effective* excess supply of labour. Accordingly, no 'signal' to the market is generated that will lead to the expansion of output and consequent reduction of unemployment (cf. also Leijonhufvud 1968, pp. 81–91). And it is the absence of such a 'signal' that Leijonhufvud (1981, ch. 6) subsequently denoted as 'effective demand failure'.

This 'failure', however – and correspondingly the failure of Walras's Law to hold in Clower's sense – is not an absolute one: for though there is no direct signal to the commodity market, an indirect one may well be generated. In particular, the very fact that the constrained equilibrium in the labour market does not represent a balancing of market forces means that the unemployed workers in this market are a potential source of a downward pressure on the money wage rate. And if this pressure is to some extent effective, the resulting decline in money wages will generate an increase in the real quantity of money, hence a decrease in the rate of interest, hence an increase in investment and consequently in aggregate demand – and this process may be reinforced by a positive real-balance effect (see chapter 19 of the *General Theory*, which, however, also emphasizes how many weak – and possibly perverse – links there are in this causal chain). Thus a sufficient condition for absolute 'effective demand failure' is the traditional classical one of absolute rigidity of money wages and prices.

An analogy (though from a completely different field) may be of help in clarifying the nature of the foregoing equilibrium in the labour market. Consider a cartel of (say) oil-producing firms, operating by means of a Central Executive for the Production of Oil (CEPO) which sets production quotas for each firm. The total quantity-constrained supply so determined, in conjunction with the demand conditions in the market, will then determine the equilibrium price for crude oil, and that equilibrium position is the relevant one for Walras's Law. But this will not be an equilibrium in the full sense of the term, for it coexists with market pressures to disturb it. In particular, the monopolistic price resulting from CEPO's policy is necessarily higher than the marginal cost of any individual member of the cartel. Hence it is to he interest of every firm in the cartel that all other firms adhere to their respective quotas and thus 'hold an umbrella' over the price, while it itself surreptitiously exceeds the quantity constraint imposed by its quota and thus moves closer to its notional supply curve as represented by its marginal-cost curve. And since in the course of time there will be some

firms who will succumb to this temptation, a temptation that increases inversely with the ratio of its quota to the total set by CEPO, actual industry output will exceed this total, with a consequent decline in the price of oil. Indeed, if such violations of cartel discipline should become widespread, its very existence would be threatened.

This analogy is, of course, not perfect. First of all, unlike workers in the labour market, the member-firms of CEPO have themselves had a voice in determining the quantity constraints. Second, and more important, any individual firm knows that by 'chiselling' and offering to sell even slightly below the cartel price, it can readily increase its sales. But analogies are never perfect: that is why they remain only analogies.

A final observation: the discussion until now has implicitly dealt with models with discrete time periods. In models with continuous time, there are two Walras's Laws: one for stocks and one for flows: one for the instantaneous planned (or constrained) purchases and sales of assets (primarily financial assets) and one for the planned (or constrained) purchases and sales of commodity flows (cf. May 1970; Foley and Sidrauski 1971, pp. 89–91; Sargent 1979, pp. 67–69; Buiter 1980). On the other hand, in a discrete-time intertemporal model, in which there exists a market for each period, there is only one Walras's Law: for in such a model, all variables have the time-dimension of a stock (see Patinkin 1972, ch. 1).

See Also

- ▶ General Equilibrium
- ▶ Temporary Equilibrium

Bibliography

Arrow, K.J. and Hahn, F.H. 1971. *General competitive analysis*. San Francisco: Holden-Day/Edinburgh: Oliver and Boyd.

Barro, R., and H.I. Grossman. 1971. A general disequilibrium model of income and employment. *American Economic Review* 61: 82–93.

Barro, R., and H.I. Grossman. 1976. *Money, employment and inflation*. Cambridge: Cambridge University Press.

Buiter, W.H. 1980. Walras's law and all that: Budget constraints and balance sheet constraints in period models and continuous time models. *International Economic Review* 21: 1–16.

Clower, R. 1965. The Keynesian counterrevolution: A theoretical appraisal. In *The theory of interest rates*, ed. F.H. Hahn and F.P.R. Brechling, 103–125. London: Macmillan.

Drazen, A. 1980. Recent developments in macroeconomic disequilibrium theory. *Econometrica* 48: 283–306.

Fitoussi, J.P. 1983. Modern macroeconomic theory: An overview. In *Modern macroeconomic theory*, ed. J.-P. Fitoussi, 1–46. Oxford: Basil Blackwell.

Foley, D.K., and M. Sidrauski. 1971. *Monetary and fiscal policy in a growing economy*. London: Macmillan.

Gale, D. 1983. *Money: In disequilibrium*. Cambridge: Cambridge University Press.

Grandmont, J.M. 1977. Temporary general equilibrium theory. *Econometrica* 45: 535–572.

Grossman, H.I. 1971. Money, interest, and prices in market disequilibrium. *Journal of Political Economy* 79: 943–961.

Hicks, J.R. 1939. *Value and capital*. Oxford: Clarendon Press.

Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan.

Lange, O. 1942. Say's Law: A restatement and criticism. In *Studies in mathematical economics and econometrics*, ed. Oscar Lange et al., 49–68. Chicago: University of Chicago Press.

Leijonhufvud, A. 1968. *On Keynesian economics and the economics of Keynes*. New York: Oxford University Press.

Leijonhufvud, A. 1981. *Information and coordination: Essays in macroeconomic theory*. Oxford: Oxford University Press.

May, J. 1970. Period analysis and continuous analysis in Patinkin's macroeconomic *model*. *Journal of Economic Theory* 2: 1–9.

Modigliani, F. 1944. Liquidity preference and the theory of interest and money. *Econometrica* 12 (January): 45–88. As reprinted in *Readings in monetary theory*, ed. F.A. Lutz and L.W. Mints, Philadelphia: Blakiston, for the American Economic Association, 1951, 186–240.

Patinkin, D. 1956. *Money, interest, and prices*. Evanston: Row, Peterson. (The material referred to appears unchanged in the second, 1965 edition.)

Patinkin, D. 1958. Liquidity preference and loanable funds: Stock and flow analysis. *Economica* 25: 300–318.

Patinkin, D. 1972. *Studies in monetary economics*. New York: Harper & Row.

Ricardo, D. 1821. *On the principles of political economy and taxation*, 3rd edn. As reprinted in *The works and correspondence of David Ricardo*, ed. P. Sraffa, vol. I. Cambridge: Cambridge University Press, 1951.

Sargent, T.J. 1979. *Macroeconomic theory*. New York: Academic Press.

Say, J.B. 1821a. *Traité d'économie politique*, 4th edn, Paris: Deterville. Trans. C.R. Prinsep under the title *A*

Treatise on Political Economy, Philadelphia, Grigg R. Elliot: 1834.
Say, J.B. 1821b. *Letters to Thomas Robert Malthus on political economy and stagnation of commerce.* London. As reprinted with an Historical Preface by H.J. Laski, London: George Harding's Bookshop, 1936; Wheeler Economic and Historical Reprints No. 2.
Walras, L. 1874. *Eléments d'économie politique pure*. Paris: Guillaumin. (Sections I–III of the work).
Walras, L. 1900. *Eléments d'économie politique pure*. 4th ed. Paris: F. Pichon.
Walras, L. 1926. *Eléments d'économie politique pure*. Definitive edition. Paris: F. Pichon (for our purposes, identical with 4th edition). Trans. William Jaffé under the title *Elements of pure economics*, London: George Allen and Unwin, 1954.

Walras's Theory of Capital

John Eatwell

The extension of Walras's (1874–7) analysis from non-capitalistic production to the case of capitalistic production involved him in the introduction of four new sets of variables: the rate of net income, i; the l prices of the capital goods, P_k, the l quantities of capital goods demanded, D_k, and the total value of saving (in terms of the numéraire), E.

The rate of net income is defined as the ratio of the rental of a unit of capital good service p_k, less charges for depreciation (μ) and insurance (v) (both of which are expressed as percentages per annum of the value of the capital good) to the value of the capital good:

$$i = \frac{p_k - (\mu + v)P_k}{P_k} \tag{1}$$

Thus the value of each capital good is equal to the net price of a unit of its service divided by the rate of net income plus charges for depreciation and insurance. In competitive equilibrium the rate of net income is equal on all capital goods, but μ and υ may differ from one capital good to another. In what follows, it will be assumed for simplicity that $\mu = \upsilon = 0$.

Walras expresses the volume of saving as the outcome of demand for an imaginary commodity (E), a unit of which yields perpetual net income at the rate i which is to be determined. In consequence the price of a unit of (E), $p_e = 1/i$. The total volume of saving derives from individuals' utility maximizing choices between units of *(E)* and units of other commodities.

Summing over individual demands for *(E)* total demand for *(E)* is

$$D_e = F_e\left(p'_k, p'_n, p'_c, i\right)$$

and hence the total value of saving, E, in terms of the numéraire is

$$E = D_e P_e = G_e\left(p'_k, p'_n, p'_c, i\right) \tag{2}$$

where p_n are rentals of non-producible factors and p_c prices of consumer goods. Thus saving is defined as a fluid homogeneous magnitude – 'savings in general'. Walras regarded the market for capital as *finance* as equivalent, in the determination of net income, to the market for capital goods. But even in his examination of the latter saving is always expressed as a quantity of value (in terms of the numéraire), whilst the quantities of particular capital goods produced are expressed in terms of their peculiar physical units. The sum of the values of new capital goods demanded equals the value of saving

$$E = P'_k D_k \tag{3}$$

Although Walras labelled as capital 'all forms of social wealth which are not used up at all or used up only after a lapse of time' the determination of the rate of net income concerns only 'mobile capitals', i.e. reproducible means of production. The essential characteristic of the stock of reproducible means of production is that its composition is determined by economic forces:

> Capital assets are destroyed and vanish, like persons; and like persons they re-appear, not, however, as a result of a natural reproduction, but as a result of economic production. (p. 217)

The essence of 'economic production' is that like all other produced commodities, capital goods

> are subject to the law of cost of production …. In equilibrium their selling prices and their cost of production are equal. (p. 171)

Thus

$$P_k = Ap_k + Mp_n \tag{4}$$

where A is the $l \times l$ matrix of input coefficients of capital goods into capital goods, and M the $l \times m$ matrix of input coefficients of non-reproducible inputs into capital goods.

Walras' complete system may be set out as follows (the lower case roman numeral identifies the equations, the preceding letter (or number) indicates the number of equations in each set):

$$\begin{aligned}
&D_c = F_c\left(p'_k, p'_n, p'_c, i\right)(n)(\mathrm{i})\\
&D_e = F_e\left(p'_k, p'_n, p'_c, i\right)(l)(\mathrm{ii})\\
&E = D_e p_e = G_e\left(p'_k, p'_n, p'_c, i\right)\\
&O_n = N'D_c + M'D_k(m)(\mathrm{iii})\\
&O_k = B'D_c + A'D_k(l)(\mathrm{iv})\\
&O_n = Q_n(m)(\mathrm{v})\\
&O_k = Q_k(l)(\mathrm{vi})\\
&p_c = Bp_k + Np_n(n)(\mathrm{vii})\\
&P_k = Ap_k + Np_n(l)(\mathrm{viii})\\
&P_k = \frac{p_k}{i}(l)(\mathrm{ix})\\
&E = P'_k D_k(1)(\mathrm{x})
\end{aligned}$$

a total of $2n + 2\ m + 4\ l + 2$ equations to determine the similar number of unknowns, (D_c, p_c, O_n, p_n, O_k, D_k, p_k, P_k, E, i). By Walras's Law one of the equations is redundant. Similarly prices are only determined up to a multiplicative constant, and hence the price of a consumption good *(A)* is set equal to 1.

Walras has introduced $2\ l + 2$ new equations ((viii), (ix), (ii) and (x)) to determine the $2\ l + 2$ additional unknowns (P_k, D_k, E, i) which are required to complete his model of competitive capitalism.

Walras clearly regards the new equations and variables as an attachment to the equations of a-capitalistic production and exchange which does not disturb the solutions of those latter equations in any significant way. Thus the determination of equilibrium conditions in the theory of capital formation is expressed solely in terms of the saving function and the total value and quantities of new capital goods produced:

> With these additional data, we have all the elements necessary for the solution of our problem. New capital goods are exchanged against the excess of income over consumption: and the condition of equality between the value of the new capital goods and the value of the excess gives us the equation required for the determination of the rate of net income and consequently for the determination of the prices of capital goods. Moreover, new capital goods are products; and the condition of equality between their selling price and their cost of production gives us the equations required for the determination of the quantities manufactured. (pp. 269–70)

A peculiarity of Walras's approach to the problem of capitalistic production is that the existence of *positive* net saving plays an important role in his analysis of the determination of the rate of net income; to the extent that he even suggests that the rate of net income can only be determined in a progressive economy (pp. 269 and 479). The rationale of this odd position will become clear as we proceed.

Walras's attempt to embed the concept of 'saving in general', and an analysis of the determination of a uniform rate of net income on the value of capital goods, within the framework he had developed in the analyses of pure exchange and of a-capitalistic production lead was to prove unsuccessful, as was demonstrated by Garegnani (1960).

The source of the problem, which is examined in the next section, may be summarized as follows. The technique Walras had developed in the preceding lessons required that the stocks of means of production, expressed in their individual physical units, should be part of the (arbitrary) data of the problem. So the rentals paid for the services of these means of production will be determined by the demands for the available stocks. In the case of produced means of production the demand-prices of new capital goods will depend on the prices of their services in relation to the rate of net income (equation (ix)). And, since currently produced capital goods are *not* available for use in the period under consideration – if they

were, the stock of capital would be unbounded! (pp. 282–3) – the demand price of any new good is determined solely by the demand for the stock of its services currently available – it will not be affected by changes in the output of that good except to the extent that it is used in its own production.

For any given set of prices each type of capital good has a particular rate of return over its cost of production. If the requirement of a uniform rate of net income is imposed on the model, the value (cost of production) of each capital good must be such as to yield that rate of return.

The cost of production of new capital goods may be altered by variations in the composition of the output of new capital goods if these variations lead to changes in the demands for stocks of services. The degree of variation in the cost of production will be determined by the differences in techniques which may be used for the production of the various capital goods, and the total value of savings which may be allocated between the capital goods. Thus, variation in the composition of capital good output will determine both the cost of production and the demand price of each new capital good, whilst the volume of saving will determine the range of that variation. Given the total value of saving the equality between the cost of production and the demand price of each capital good is the condition of equilibrium, and variation in the composition of the output of new capital goods is the only means whereby this condition may be satisfied. Thus this composition must not be fixed by any condition outside the system (i)–(x). For example, if demand functions for capital goods (as functions of all prices and i) were added to the system it would be *overdetermined*, for this would involve adding extra equations to the system without the addition of unknowns. Similarly, if net savings were zero, the composition of output would be determined by the requirements of replacement, and the equations are, once again, overdetermined. Walras, by confining the analysis to an economy in which net saving is positive, obscures this difficulty. Walras's system can only admit of 'saving in general' as demand for the total value of new capital goods produced. It cannot accommodate demand functions for individual capital goods. Even so, the system is generally inconsistent.

Existence of a Solution to the Equations of Capital Formation

The conditions which must be satisfied if there is to be a solution to Walras's equations of capital formation, and the rate of net income determined, will now be investigated under a simplifying – but none the less general – assumption.

The endowment of the economy is assumed to consist solely of stocks of reproducible means of production, these stocks being of arbitrary size. Only one technique of production is available to produce each commodity. Thus the equation system (i)–(x) above must be modified by the elimination of all reference to prices or quantities of non-produced means of production.

Since i is uniform, from equations (viii) and (ix) this rate must be such that

$$Ap_k i = p_k \tag{5}$$

and hence

$$0 = [I\lambda - A]p_k \tag{6}$$

where $\lambda = 1/i$. By the Frobenius theorem it is known that since A is non-negative and (we presume) indecomposable, there is only one value of the characteristic roots λ_i with which a positive eigenvector p_k is associated. All other λ_i are associated with vectors which contain negative prices and are therefore economically meaningless. Thus there can be only one value of i, and only one vector of rentals of the services of capital goods, p_k, consistent with the existence of a uniform i on all produced means of production. Once i and p_k are determined, then by equation (vii) p_c are also determined, as are D_c, E and P_k by equations (i), (ii) and (ix).

Only the demands for capital goods remain to be determined, and only the equations expressing the equality between the endowment of capital goods and the demand for new capital goods

(iv) and the equation expressing equality between the value of gross saving and the value of new capital goods produced (x) remain to be satisfied. It should be noted that in equilibrium the demand for each stock of capital goods (i.e. for each stock of capital goods services) must be equal to the size of the stock, since all p_k are positive.

By Walras's Law

$$p_c' D_c + E = p_k' Q_k \tag{7}$$

i.e. the amount spent (measured in terms of the numéraire) on consumption goods, plus the amount saved, must be equal to the income earned on the endowment of means of production. Since, by equations (vii) consumption goods' prices are equal in equilibrium to their costs of production, then the total value of the output of consumption goods is equal to the value of the services used in their production.

$$p_c' D_c = p_k' Q_k^c \tag{8}$$

where Q_k^c denotes the vector of quantities of means of production used in the production of consumption goods. Hence

$$E = p_k' Q_k - p_k' Q_k^c = p_k' Q_k^a \tag{9}$$

the value of saving is equal to the total value of the services yielded by the amounts of the initial endowments *available* for the construction of new capital goods (Q_k^a), once the requirements of consumption goods production have been deducted from the original stocks.

The stock of capital–good services *required* for the construction of new capital goods (Q_k^i) may be determined from the two conditions:

$$E = P_k' D_k \text{ and } Q_k^i = A' D_k,$$

Walras's equations are consistent only if $Q_k^i = Q_k^a$; that is if there exists a vector of demands for new capital goods D_k such that this latter condition is satisfied. All elements of D_k must be nonnegative, and, in turn, all input requirements are non-negative;

$$D_k = (A')^{-1} Q_k^i > 0; Q_k^i \geq 0 \tag{10}$$

However, there is nothing in the specification of the model to ensure that $Q_k^a \geq 0$. The original endowment Q_k must be non-negative. But the solution may imply that the amount of any one element of the endowment available for the production of new capital goods is negative, to the extent of

$$-\sum_c b_{jc} D_c,$$

the quantity of capital good j used in the production of consumption goods. Indeed, only one element of Q_k^a need be positive (this is essential if E is to be positive). Thus the set Q_k^a is bounded from below by the condition $Q_k^a \geq -B' D_c$. So if any element of Q_k^a is less than zero, a condition that is as likely as its converse, then the data and the equations are unequivocally inconsistent. Consistency can be attained only by chance. This is a sufficient criticism of Walras' system.

Moreover, even if Q_k^a should be non-negative, condition (10) may render the system inconsistent. The vectors Q_k^i which satisfy (10) lie in a convex cone within the positive orthant, for these vectors must be non-negative linear combinations of the columns of A'. In general, this cone will not be the entire positive orthant – this would only be the case when each productive service appeared only once as the sole input into the production of a single capital good. Hence, in general, some $Q_k^a \geq 0$ cannot be equated with a vector Q_k^i which satisfies (10).

This analysis is unaffected if it is assumed that there is more than one technique available for the production of each commodity. The technique which will be chosen is that which yields the highest rate of net income. This technique will be used whatever may be the composition of demand or of the endowment.

The rationale of this result is that there is no reason to suppose that prices paid for the services of the stocks of reproducible means of production will correspond to those prices which would result in a uniform rate of net income, the condition of long-run equilibrium.

Including non-produced means of production in the analysis will not render the model consistent. The vector of rentals which clears the markets for non-produced factors will have associated with it a vector of rentals for capital goods which would be compatible with a uniform rate of net income. Except by a fluke, this vector of capital goods rentals will not be market clearing.

An Alternative Solution

Walras himself became aware, in the 4th, definitive, edition of the *Elements*, that his equations of capital formation might not admit of a solution:

> If we suppose that all fixed capital goods proper . . . are already found in the economy in quantities Q_k . . . and that their gross and net incomes are paid for at prices determined by the system of production equations and by the rates of depreciation and insurance, it is not at all certain that the amount of savings E will be adequate for the manufacture of new fixed capital goods proper in just such quantities as will satisfy the last l equations of the above system (p. 308).

His solution was, however, an evasion of the real problem:

> On the other hand, in an economy in normal operation which has only to maintain itself in equilibrium, we may suppose the last equations to be satisfied (p. 308).

i.e. being in equilibrium we may hope for the attainment of equilibrium!

> But Walras also recommended a way out of the dilemma
>
> All we could be sure of, under these circumstances of insufficient savings to ensure equilibrium is (1) that the utility of new capital goods would be maximized if the first new capital goods to be manufactured were those yielding the highest rate of net income, and (2) that this is precisely the order in which new capital goods would be manufactured under a system of free competition (p. 308).

If we take up this proposal the equations (viii) in Walras' system should be modified, the equality between cost of production and demand price being replaced by the inequality:

$$P_k \leq Ap_k + Mp_n \qquad \text{(viii}'\text{)}$$

i.e. demand price is less than or equal to cost of production – with the proviso that in the cases of those capital goods for which the inequality holds output will be zero.

The important role which the replacement of equalities by inequalities plays in the establishment of an economically meaningful solution to the equations of exchange and of acapitalistic production and exchange is well known. In the case of consumption goods a demand price which is less than the cost of production for any positive output means that output of that good must be zero. In the case of non-producible inputs an endowment which is greater than the quantity demanded at any positive price results in the price of that input being zero. Both these circumstances have a clear economic meaning. Reproducible means of production share characteristics of both and the economic meaning of the use of the inequality is less clear. The services of the stock of a reproducible input may command a positive price and yet new units will not be produced for the demand price of that commodity is less than its cost of production. The rate of net income earned in the production of such a commodity, calculated at the ruling prices, will be less than the ruling rate. But a situation in which the rate of net income is not the same on all produced means of production does not conform to the conventional notion of longrun equilibrium. An 'equilibrium' defined with inequalities between demand price and cost of production of some capital goods is a curious hybrid, for although the prices of all non-produced means of production are uniform throughout the economy (a long-run equilibrium condition), the rate of net income is not uniform (a short-run condition). This hybridization cannot be justified by considerations affecting the relative mobility of resources in the two classifications, since mobility of non-reproducible inputs between uses is customarily attained by changes in the structure of the stock of producible inputs with which they are combined. This situation arises not from any particular view of the actual operation of a capitalistic economy but is dictated by the necessity, in Walras' theory, of expressing the stock of reproducible means of production as a set of arbitrary (physical) magnitudes.

In a Note to Chapter III of his *Equilibrium, Stability and Growth*, Michio Morishima has adopted Walras's strategem and demonstrated (with the aid of some additional assumptions) that a solution to the modified system exists. His proof follows the mathematical procedures developed by Wald, and Arrow and Debreu, adapted to the case of capital formation. The technique developed above in the analysis of Walras's equations of capital formation will be used to investigate Morishima's analysis.

The investigation will be conducted under the assumption that all means of production are reproducible, and that only one technique is available for the production of each produced commodity (consumption goods and capital goods). At least one input coefficient is positive for any output. These assumptions are made solely in the interests of simplicity. It will be evident from what follows that the argument could readily be extended to include non-reproducible means of production and many possible techniques.

Suppose that all capital goods are produced and that the price system is that which corresponds to a uniform rate of net income. Then, for the reasons outlined above there will in general be a discrepancy between the stocks of capital goods' services available for the production of new capital goods, and the set of possible combinations of stocks required if all saving is to be absorbed and the outputs of new machines are to be non-negative. Some elements of the difference $Q_k^a - Q_k^i$ (for any Q_k^i selected from the set of possible alternatives) will be positive and others negative. Those elements which are positive relate to an endowment which is greater than that which is absorbed in the production of both consumption goods and capital goods at existing prices. Those elements which are negative relate to an endowment smaller than demand for its services at existing prices.

To attempt to remove the discrepancies select the good for which ich the ratio $\left(q_j^a - q_i^i\right)/q_j$ is greatest, and set the output of the good at zero. The price of its services (which will be denoted by π_j) may now be set anywhere in the range between zero and its price when produced (p_j). The inequality in (viii)′ will therefore hold for good j. Since there was previously an excess supply of this good then it might be expected that π_j set lower than p_j would tend to reduce that excess by encouraging substitution in consumption toward j intensive goods. Such changes may also supplement the available stocks of those services for which there was previously an excess demand. Any value of $\pi_j < p_j$ will result in a higher rate of net income implicit in the price equations of those capital goods which are produced; this may, in turn, tend to increase saving and hence available stocks of all capital goods for capital good production. As good j is not produced it acquires the role in price system of a nonproducible means of production, and the analysis may be pursued once more in the manner outlined above; i.e. by relating the intersection of the set $Q_k^i (k \neq j)$, which for all possible pairs i, π_j ($\pi_j < p_j$) results in the absorption of saving, with the set of Q_k^j which satisfy the condition

$$q_j^a \geq q_j^i$$

(if the inequality holds $\pi_j = 0$), to the set Q_k^a defined by

$$E - \pi_j q_j^a = p_k' Q_k^a (k \neq j)$$

There is no reason to believe that the discrepancy will disappear, and so the procedure may be repeated; the production of the capital good for which the value of $\left(q_k^a - q_k^i\right)/q_k$ is the greatest should be set equal to zero, and π_k set below p_k. The analysis is then conducted as if there were two non-produced inputs.

The quantity available for capital good production of a capital good previously eliminated from production may fall below zero for all $\pi_k < p_j$ as the structure of production changes with successive eliminations (p_j being calculated as the cost of production of good j at the prices ruling at the stage of the analysis under consideration, not at the stage at which it was eliminated). In such circumstances the price of j must be raised to p_j, which means that it is reintroduced into production, a different good is selected for elimination,

and the process continues in the same manner as before.

At each stage of the process of elimination it may happen that there is a set of Q_k^i (of those capital goods still in production), appropriate to the prices associated with a uniform rate of net income on the goods produced. But this event is similar to the attainment of a uniform rate of net income when all capital goods are produced, that is, a configuration of the endowment which happens to result in a Q_k^a which falls in the subset Q_k^i. Since the initial endowment is arbitrary it cannot be claimed that in such a case the existence of an equilibrium to the equations has been proven – since in general, for all circumstances, it has not. A configuration of the endowment can always be found which would result in the elimination being taken a stage further. Each stage of the elimination merely recreates this situation, and thus there is no reason to believe that the process will cease.

The process of elimination may thus continue until only two capital goods remain in production, and $l - 2$ capital goods are consigned to the category of non-produced means of production. There is still no necessity that Q_k^a should equal a feasible $Q(k$ not including the $l - 2$ eliminated goods). That good which is in excess supply should be eliminated. The system now only contains *one* produced means of production, and the savings function (ii) is, in effect, the demand function for that good alone. No constraints are now imposed on the configuration of prices by the condition that the rate of net income on the produced means should be uniform – for since only one capital good is produced there can only be one value of the rate of net income.

The only case in which there must necessarily be a solution to the system is that in which only one capital good is produced. Condition (viii)′ will hold with just one equality.

Thus, apart from the chance case in which the elimination process is halted with more than one capital-good in production, a maximum uniform rate of net income is attained only when just one capital-good is produced. The rate of net income defined in the production of the single good produced is used to capitalize the value of non-produced capital goods and hence these 'earn' the rate of net income by definition. Morishima's model is thus yet another example of the use in neoclassical models of the 'one-produced input world' assumption, input is to be the one produced is endogenous to the model.

Walras' analysis of capital formation and credit, far from being the triumphant confirmation of his theory of pure economics, is a failure which brings his whole system into question. He is unable to overcome the contradiction between saving in general as a homogeneous fluid magnitude and the heterogeneity of capital goods. This contradiction could be overcome by expressing the endowment of capital goods as a single magnitude – their value. But the value of the endowment cannot be part of the data of the problem without engendering circular reasoning. Walras, in avoiding this circularity, constructed a system in which whilst the method of specifying the data is logically sound, the equations are inconsistent.

See Also

- ▶ Arrow–Debreu Model of General Equilibrium
- ▶ General Equilibrium

Bibliography

Garegnani, P. 1960. *Il capitable nelle teoria della distribuzione*. Milan: Guiffre.

Morishima, M. 1964. *Equilibrium, stability and growth*. Oxford: Clarendon Press.

Walras, L. 1874–7. *Eléments d'économie politique pure*. Trans. and ed. W. Jaffé as *Elements of Pure Economics*, London: Allen & Unwin, 1954.

Wants

Mary Douglas

To be in want is not to have. The obverse of want is satisfaction or having the wherewithal for happiness. Much Eastern philosophy recommends

happiness based on few wants, just as much of Western philosophical comment condemns excessive wants. The economists' view is different. They tend to worry when an economy comes to rest at a low level of wants and to feel more sanguine when the demand for new possessions goes up, even if they become worried again if demand is inflationary. They are clearly interested in wants. Yet the way that demand for goods is treated within economic theory blocks their curiosity about how wants are generated. This is not to say that distinguished economists have not seriously pondered the subject. Many have produced catalogues of wants, sometimes contrasting material with spiritual satisfications, sometimes comparing long-term with short-term wants, or psychic joys (such as music or affection) with physical requirements (such as food and warmth). Such lists tend to dangle free of theoretical constraints. They remain mere lists whose parts do not mesh into any theory.

Anthropology is in no state to supplement this missing element in economics. Both disciplines have an explicit theory about the circulation of goods but only an implicit theory of wants. In economics the implicit assumption is that the origin of wants is to be found inside the individual's physical and psychic constitution. In anthropology, the implicit assumption is that wants are defined and standardized in social interaction. This latter view makes a better start for thinking about wants because it integrates the choices of the individual agent within a model of the whole economy, whereas economics leaves the choices unexplained except in regard to price. To get into such a starting position economics would need to modify the concept of the consumer as an independent rational agent choosing to satisfy personal needs. It would also need to take an interest in what happens to goods after purchase. The word consumption implies that the goods are destined to be used up in the purchaser's home. Once in the shopping basket they hold little interest for economic theory, but that is the point at which the anthropologist's interest begins. Most goods are likely to be widely shared or passed from hand to hand over a certain span of time. Instead of someone who buys for private purposes the consumer would have to be seen as someone engaged in long-term interactions with other social beings and using goods to promote the particular social patterns that he values.

For the anthropologist, wants are primarily generated in social life; if this is so, when the pace of social interaction slackens, demand for possessions will go down. This approach began with Malinowski's account (1922) of Trobriand Islanders going in canoes to exchange shell ornaments and other products through vast reaches of the Pacific. These people made a clear distinction between trade and gift, and used both to build up partnerships which were not only profitable but supported their intentions within their local political systems. Marcel Mauss (1925) extended these insights to a general theory of solidarity based on reciprocal obligation. From these beginnings, succeeding generations of anthropologists came to study all kinds of transfers of rights and property as flows marking the important channels of social obligation. The focus on types of reciprocity as the basis of solidarity was formalized by Claude Lévi-Strauss in a general theory of kinship. One kind of repeated marriage patterns can produce long lines of exchange embracing everyone in the community and all generations in a generalized system of transfers; another has more restricted effects, linking only two or three descent lines; endogamy is the limit case of marrying-in at the expense of a wider solidarity. Such variations have direct implication for the political system and for the economy. Marrying or procreating appear as part of the total system of reproduction. It has generally been assumed that this kind of analysis applies only to societies in which market organization is weakly developed. However, it can be argued that the sharp disjunction between market and non-market is an artifact of economic theory and one which makes theorizing about demand peculiarly difficult.

The implicit assumption in anthropology is that individual wants are standardized by the same processes that establish social solidarity. Put crudely, the reason anyone wants anything (physical needs apart) is for sharing with or showing or giving to someone else in recognition of similar gestures, gifts or services received in the

past. On this assumption, being severely in want means being unable to take part in the major reciprocal exchanges by which future entitlements are conferred. This is no trivial matter. Lacking entitlement is equivalent to becoming a thirdclass citizen or even to losing civic status. Anyone who exerts no claims on the rest of society finds that his sons and daughters are not sought in marriage; he wants for protection and can expect an indigent old age. Such a theory of wants is capable of being made explicit and generalized beyond the range of societies the anthropologists usually study. It would enable economic theory to integrate social life, family structure, demography and the labour market into the rest of the economy. The obstacle lies in the way that the theory of demand has been formulated.

The original utilitarian philosophy presupposed that wants are in some sense commensurable. Mathematical treatments of wants based on this assumption were already being applied to economic analysis when the theory of diminishing marginal utility was worked out independently in 1871 by Carl Menger and W.S. Jevons; Walras also arrived at it in the same year and independently, though he published a little later. Such a simultaneous convergence upon an intricate idea would be quite impossible if the common infrastructure of theory was not already in place. The relevant point for an article on wants is that the problem to which they all found the same answer was not how to formulate a theory of wants, not at all. The problem was how to formulate the concept of demand so as to harmonize this part of economic theory with the rest of the theory of supply and demand. Diminishing marginal utility means that an individual purchaser gets marginally less satisfaction from each additional increment of a commodity. The underlying metaphor is physical: more and more bread or beer or beef give less benefit to the eater and bigger and bigger doses of a medicine may actually harm instead of curing the patient. By incorporating diminishing marginal satisfaction for the consumer, demand theory matches the theory of supply according to which marginal costs increase with increase in the volume of production. Beyond a certain point, rising costs mean that the price must rise to encourage extra output. As the marginal utility to the consumer falls, he becomes less willing to spend his income on it. The rising supply curve cuts the falling demand curve and the see-saw comes to rest.

Whereas the theories of production, exchange and capital formation drawn up on this model only had to face technical criticism, when the model was applied to wants, philosophical and political objections appeared. How can human wants be given numerical expression? How can one person's wants be compared with another's? How can such comparisons not carry a load of political prejudice?

In the history of science it often happens that a theory does not apply well to the behaviour it is supposed to explain, because its coherence within a larger theory prevents the bad fit with data being taken seriously. In this case the theory of demand cannot give an account of wants simply because this is not what it was designed for. The very completeness of its embedding in the larger, unified theory makes it incapable of focusing on its nominal subject matter. It gives a gravely misleading account of wants for the following reasons.

First, violence is done to the concept of the individual consumer by making it parallel to the concept of the individual firm. The consumer's wants do not correspond to the profit maximizing objectives of the firm. This is essentially because the consumer is not an individual among other consumers as the firm is an individual in the market. In order to live in a society the individual consumer has to develop categories of thought and tastes conformable with those of his fellows. The processes of standardization which should be at the centre of a theory of wants are ignored by economic theory. In default of a theory of how wants are collectively generated, it falls back on hidden assumptions about the priority of physical needs. As a result of this heavy disadvantage in thinking about wants, the threat of famine tends to be perceived as a physical failure of the supply of physical necessities, not as a failure of demand. It is true that in a famine the would-be buyers have nothing to offer in exchange for the food they need. But to know how they got into that situation is to see how demand is generated by a variety of reciprocal exchanges which guarantee future

entitlements. A.K. Sen (1981) has argued that the misdiagnosis of the causes of major famines is due to inability to see how individuals enter the economic system and stay in it. Without what he calls exchange entitlements, individuals and their dependents are vulnerable to shocks in the economic system. Such a systemic view of the way that wants enter the economy and are shaped by social and legal processes is necessary if the anthropological approach is to be joined with economics in a general theory of wants. In this perspective the pattern of wants is the surface appearance of a pattern of social relations and social opportunities. Goods are needed as aids to interaction and as clues for constructing intelligible worlds. The consumer is engaged in a continual task of grading goods and occasions and matching them appropriately, as every market researcher knows. It should be useful for a theory of demand to take the social pressures into account. The more isolation and segregation, the more is demand dampened, the more the interaction, the more the need for a symbolic system articulated by finely graded patterns of consumption.

Third, the theory makes one connection (price) between consumption and production but misses another. It treats tastes as personal and subjective and so uninfluenced by the organization of work. But tastes depend upon shared consumption, so the timing of work, the location of homes, the life cycle expectations which are engendered by different occupations, all these and other aspects of the labour market influence the standardization of wants.

To correct these weaknesses in the only theory that claims to be a theory of wants would involve taking much more interest in shared cultural categories that characterize a community. Economists expect to apply their theories to public policy. But whenever they are tempted to speak of what is good for a community, their theory leads to contradiction. As Arrow's theorem proves, the ranked preferences of several individuals cannot necessarily be aggregated into a single ordered set for them all unless, of course, they happen to have the same preferences. In respect of material things they very frequently do. But there is no theory about how this comes to pass. So the theory is at a loss when it comes to thinking about community welfare. Starting from incommensurable, subjective, individual preferences it cannot proceed to theorize about what a community wants. Yet, there seems to be no inherent reason why a theory of wants, which gives credit to their social origins and their social definition and to their community-imposed character, should not serve the needs of economic theory as well as, better than, the one which has historically developed from the concept of the individual as a surrogate for the firm.

See Also

- ▶ Economic Anthropology
- ▶ Social Choice

Bibliography

Douglas, M., and B.C. Isherwood. 1979. *The world of goods*. New York: Basic Books.

Edgeworth, F.Y. 1881. *Mathematical psychics. An essay on the application of mathematics to the moral sciences*. London: Kegan Paul.

Lévi-Strauss, C. 1949. *Les structures élémentaires de la parenté*. Paris: Presses Universitaires.

Mackenzie, D. 1981. *Statistics in Britain, 1865–1930*. Edinburgh: Edinburgh University Press.

Malinowski, B. 1922. *Argonauts of the Western Pacific*. London: Routledge & Kegan Paul.

Mauss, M. 1925. Essai sur le don. *L'année sociologique*, 2nd Series, vol. 1, 23–4. Trans. as *The Gift*, London: Cohen and West.

Sen, A.K. 1981. *Poverty and famines: An essay on entitlement and deprivation*. Oxford: Clarendon Press.

War and Economics

Craufurd D. Goodwin

Abstract

War has always intrigued economic thinkers: how economic interdependence affects the likelihood of conflict, the costs and benefits of aggression, how to fight a successful war and achieve a stable peace. Many prominent early economists had useful things to say, for

example Smith, J.B. Clark, Pigou, Veblen and J.M. Keynes. Economists came into their element in the Second World War and the Cold War; welcomed into government, they distinguished themselves as macroeconomists, microeconomists, broad-based advisers, public administrators, and public figures. By the 1990s economists concerned with war and peace had established their own subdiscipline, with journals, textbooks and college courses.

Keywords

Angell, N.; Autarky; Baran, P. A.; Cassel, G.; Clark, J. B.; Clark, J. M.; Classical economics; Colm, G.; Colonialism; Council of Economic Advisers (USA); Currie, L.; Defence economics; Deterrence; Enlightenment; Galbraith, J. K.; Game theory; Heterodox economics; Hobson, J. A.; Keynes, J. M.; Kindleberger, C.; Lloyd, H.; Malthus, T. R.; Marginal revolution; Mill, J. S.; Operations research; Parsons, T.; Pigou, A. C.; Rostow, W. W.; Rule of law; Schelling, T.; Smith, A.; Sweezy, P. M.; Veblen, T.; War and economics

JEL Classifications

B2

Prehistory

Political economy emerged from moral philosophy as a distinct field of study in the 18th century. Before that economic ideas were part of a continuum developing from the speculation of Greek philosophers about how to achieve the good life, the advice from medieval schoolmen about how to achieve salvation, and the recommendations of 17th century merchants and public administrators about how to strengthen the emerging nation-states. In this early literature economic ideas were entwined with concerns about communal security. No matter whether the unit of concern was a nation, a city-state, or a manor, conquest by an external enemy was seen as a catastrophe. Conversely, great economic benefit was seen to come from military victory.

Hesiod in eighth-century BC Greece suggested that economic isolation and selfsufficiency might be the best protection against war because for voracious enemies out-of-sight was out-of-mind. Indeed, the ideal communities imagined by Plato in Atlantis, Sir Thomas More in Utopia, and Francis Bacon in New Atlantis were all remote from their neighbours (Lowry 1991, p. 6). Security, they implied, lay in isolation and economic autarky. In economic interdependence there was danger. Economic interchange requires human interaction and from such contacts emerge frictions, cupidity, and ultimately physical violence. The state dealt with domestic violence through the police and the courts. Correspondingly international trade and flows of resources across borders should be constrained by the state to reduce tensions that could lead to war. The ancients recognized that, if war should begin, the structure of the economy could profoundly affect the capacity of a state to defend itself. The early Greeks appreciated that the highly efficient hoplite military formation, as well as other military tactics, depended upon the strength and high morale of their small-scale farming population. As an effective fighting machine a labour force of docile day labourers just would not do. Similarly, the citizen-based merchant marine of Athens was the foundation upon which her powerful navy rested (Lowry 1991, p. 7).

Warfare was treated by the Greek philosophers and their admirers in the Middle Ages as simply one among many possible economic activities. Trade, colonization and conquest of one's neighbours were alternative uses of national resources. The task of leaders was to identify the best among them. This calculation required the kind of hard-headed cost–benefit analysis that became the hallmark of later economics. Erasmus wrote in 1516 in a tract for the future Charles V of Spain: 'when the prince has put away all personal feelings, let him take a rational estimate long enough to reckon what the war will cost and whether the final end to be gained is worth that much – even if victory is certain' (Lowry 1991, p. 17). Erasmus warned that any responsible leader must guard against the self-interested advice of private persons who might gain from war while the community at large would lose. 'It too often happens that nobles, who

are more lavish than their means allow, when the opportunity is presented stir up war in order to replenish their resources even at home by the plunder of their peoples' (1991, p. 17). This advice sounds like an early warning against a military–industrial complex!

War and Classical Political Economy

To a degree, the classical political economy that appeared in the 18th century was an extension of the earlier 'mirror for princes' literature to which Erasmus contributed. It, too, was intended to guide statesmen in the preparation and implementation of public policy. An important difference, however, was that the political economists thought that they brought much greater impartiality and detachment to their work. Policy implications, they believed, emerged ineluctably from their theory, and ad hoc reasoning could be replaced by principle (Silberner 1946). Adam Smith took the same cost–benefit stance towards warfare as Erasmus, but he developed its implications more fully, and he drew from them optimistic prospects for the future (Smith 1776, pp. 689–708; Goodwin 1991, pp. 24–8). In the tradition of Enlightenment thinking, Smith presumed good political leaders would be rational, and the challenge, therefore, was to construct circumstances in which they would select peace over war. Happily, Smith observed, conditions were likely to emerge from growth in 'the necessaries, comforts and conveniences of life' that would make nations ever more peaceful. A consequence of economic growth was that, over time, the opportunity costs of warfare rose – reflecting both the direct costs of munitions and supplies and the production forgone from military servicemen in the prime of their productive lives. On the other hand, the value of potential spoils of war to an advancing country would no more than remain constant. As the costs rose and the potential gains remained fixed, the likelihood of benefiting from aggression declined, and the warlike tendencies of rational nations fell steadily as they grew economically. The military preparedness of a prosperous nation, therefore, would seldom be for offence. However, preparedness remained critical for defence, especially against less prosperous neighbours whose cost–benefit calculations of aggression against them might for some time remain positive. The main task of the military establishment in any prosperous nation was to make the costs of aggression against it higher than any plausible gain. In this way the notion of rational deterrence was given a precise exposition by Smith.

An intriguing yet minor Enlightenment thinker who proposed an economic doctrine while applying it also to military strategy and tactics was the Welshman Henry Lloyd, a major-general and professional mercenary in the wars of the 18th century, parodied by Gilbert and Sullivan in *Pirates of Penzance* for his commitment to social theory (Speelman 2002).

Thomas Robert Malthus among the classical economists provided a reason rooted in demographic growth to explain why the costs of conflict should always be kept clearly before potential aggressors (Malthus 1826, p. 47; Goodwin 1991, p. 26). When an increasing population pressed hard upon natural resources a suffering nation was likely to look abroad for relief and conflict would ensue. Indeed, the slaughter of warfare became one of the powerful 'positive' checks to population growth. Deterrence costs should always be kept sufficient to make certain that aggression would be carried out by overpopulated countries against someone else. John Stuart Mill concluded in 1848 that deterrence had by that date been so successful that war had become likely only in colonies where 'savages' prepared unreasonable cost-benefit calculations (Mill 1848, p. 707).

Economists, beginning with Smith, were quick to see an analogy between the security needs of all nations taken together and those of small communities that formed themselves into nation-states to benefit from the rule of law. Costs of protection could be reduced globally through international agreements. Smith favoured for Britain a kind of 'commonwealth' that would include former colonies. Later economists were enthusiastic supporters of such institutions of world government as the International Court of Justice at The Hague, the League of Nations and the United Nations.

Economic thinkers have often puzzled about just what is required to fight a successful war.

Some pamphleteers in the 17th century thought that a stock of 'treasure' was needed. Smith thought, to the contrary, that a satisfactory flow of product was necessary; a vibrant and robust economy would allow the state to 'draw from their subjects extraordinary aids upon extraordinary occasions' (Smith 1776, p. 446). Costs to the economy, he thought, were necessarily different in a small war from a large one. For example, soldiers who were not seriously challenged by the prospect of a slight skirmish in a minor war might be persuaded to volunteer at wages below market rates, led on by their 'youthful fancies' and 'romantic lapses'. The taxpayers in this situation, moreover, might not complain about modest levies because 'this amusement compensates the small difference between the taxes which they pay on account of the war, and those which they are accustomed to pay in time of peace' (1776, p. 920). In a full-scale war, on the other hand, soldiers fearing personal injury demanded payment at top market rates, taxpayers balked at the levies upon them, and it became necessary to issue public debt as the only way to raise revenue. Happily the net costs of war usually were low because productive capital was seldom damaged and, as the classical economist John Rae observed, the turmoil of conflict seemed to stimulate the powers of invention (Rae 1834, pp. 222–3).

Marginal Economics Applied to War

An important turn in the evolution of modern economics occurred in the 1870s when the mainstream accepted utility maximization as its norm and incremental analysis as its method (Black et al. 1973, p. 19). This shift took place in the middle of the Pax Britannica, one of the most peaceful periods in European history, and the new analytical tools had seldom to be used to explain warfare. Under the marginal paradigm all destructive conflict was treated as socially irrational and contributing to net social disutility. War was mentioned in relation to public finance where it was usually categorized as a 'bad', like crime and disease, the opposite of a 'good'. The main questions were how many public resources should be allocated to deterrence and how should these resources be deployed. Two writers who epitomized the marginalist approach to conflict in the years leading up to the First World War were the English economist/journalist Norman Angell, and the pioneering American price theorist John Bates Clark (Barber 1991a, pp. 61–84). Angell used economic reasoning in a book on war entitled *The Great Illusion* (1910), which reached a large and influential readership. He claimed that because of widespread international economic integration war had become everywhere futile and irrational. No one could gain from a fight. The economic dependence of virtually all countries on their neighbours for markets, products and materials made it impossible for them to achieve a net gain through conquest. This was not to say that war was impossible, only that it was unlikely by the economist's logic. This position, of course, was the exact opposite of the venerable idea developed by Plato that economic integration was dangerous and economic autarky could be the road to world peace. In Angell's view, Adam Smith's desired condition for deterrence had been achieved in which all powerful nations that made a disinterested study must conclude that the costs of war exceeded the benefits. Even indemnity payments to winners – if they could be enforced, as they had been after the Franco–Prussian War – were likely to cause serious market dislocations in the recipient nation as well as in the payer and were of little net gain to anyone.

John Bates Clark of Columbia University, the most prominent American marginal economist, studied war as part-time Director of the Division of Economics and History of the Carnegie Endowment for International Peace, founded by Andrew Carnegie in 1910. In 1911 Clark mobilized in Bern most of the celebrated economists of the time from around the world to explore how economics could help achieve permanent world peace. Those in attendance were conscious of the current threat of war and agreed to turn their talents to how it could be avoided. Clark's hope was to identify adjustments that might be made to incentives and rewards in the global economy to reduce the likelihood of conflict. There was a prevailing sense

among the economists that tensions grew out of structural errors and misunderstandings, and in this Clark saw an analogy with industrial relations. Mediation and arbitration that worked in one sphere might work also in the other. One task for economists could be to propose equitable middle ground between potential combatants.

Some Dissenting Economists' Voices

Despite Clark's and Angell's optimism about the capacity of economics to show the way to a stable world peace, heterodox economists in the socialist, historical and emerging American Institutionalist traditions continued to worry that economic forces within the industrial nations would lead inevitably to conflict. Thorstein Veblen, John A. Hobson and others warned of tensions based on class division and the proclivity of one social group to exploit another (Biddle and Samuels 1991; Davis 1991). When these tensions spread overseas through colonial expansion and an aggressive search for profits the possibility of international conflict became serious.

No matter whether economists writing about conflict in the years before the First World War were in the respectable mainstream or in the professional underworld, their reflections were mainly those of outsiders. War was a subject about which their theory seemed to have something to say, and so they said it. But there is little evidence that their observations had a measurable impact on the thought or actions of those who decided upon peace and war.

The First World War threw many prominent economists nearly into despair: the war was so costly, so barbarous, and so irrational! It seemed even to cast doubt on the controlling force of reason in society. Francis Hirst, editor of The *Economist,* in a personal manifesto in 1915 entitled 'The Political Economy of War' saw ahead only 'social and economic ruin' (Barber 1991a, p. 72). Gustav Cassel in Sweden and the young John Maynard Keynes in Britain were deeply discouraged to find that despite the global integration that Angell and others had described the belligerents seemed able and willing to continue the conflict for a prolonged period, far after any rational leaders would have been expected to make peace (Barber 1991a, pp. 76–9). After the war finally came to a close Keynes wrote in his celebrated polemic against the peace treaty, *The Economic Consequences of the Peace* (1920), that sheer vindictiveness had overcome good sense. Later, Keynes, John Bates Clark, and other economists were outraged that the American Senate refused to ratify American participation in the League of Nations. It seemed to them that once again, prejudice and political opportunism had overwhelmed reason on vital matters related to peace and war. So what use was economics in the search for peace? After a century of giving positive answers to this question economists had now to admit 'perhaps not very much'!

Economists in the years between the two world wars, apparently humbled by the seeming irrelevancy of their efforts in the years surrounding the first war, were generally reluctant to engage with subjects related to peace, war and national security. They had lost their self-confidence and they responded by withdrawing from the field. War for most economists simply was no longer a respectable area for attention. It became, instead, a condition exogenous to their models and left to be analysed by other disciplines. Symbolically, the Carnegie Endowment for International Peace restructured its efforts in search of peace away from the abortive multi-national study of the economics of war to a 132-volume economic and social history of the First World War under the direction of the distinguished Columbia historian James T. Shotwell. The clear message of the Carnegie switch was that since the deductive methods of the marginal economists had not shown the way to peace it was time to give inductive historians a chance.

One of the few exceptions to the interwar inattention by economists to peace and war came from an unlikely source, the Cambridge theorist and successor to Alfred Marshall, A.C. Pigou. In his *Political Economy of War* (1921) Pigou set out to shed light on 'the anatomy and physiology' of 'a strained and stressed economy' (Barber 1991b, p. 131). Another exception was John Maurice Clark who, with his father John Bates Clark,

attempted to estimate the full economic cost of the First World War, including the value of lost human as well as physical capital (Clark 1931).

It appeared to many professional economists in the years between the wars that the dragon they had to slay in order to achieve peace was not in fact war itself but rather the conditions that led to war – at that time, economic depression and unemployment. So long as workers could not find jobs, and bankruptcies destroyed entrepreneurs, civil insurrection was likely. And when economic distress led to the rise of demagogues then world peace was truly at risk. Enquiry into the root causes of macroeconomic crisis proceeded at several levels. In Europe, business cycle institutes looked for disturbing patterns in worldwide economic activity and for means of coping cooperatively with distress (De Marchi 1991). In the United States, economists inside and out of the Roosevelt administration experimented with a variety of new stimulative and regulatory mechanisms. In Britain, Keynes and his students at the University of Cambridge developed and propagated a new macroeconomics focused on the adequacy of aggregate effective demand, and they pointed to fiscal policy as the means to achieve economic stability. Keynes recalled T.R. Malthus's observation that wars of aggression could emerge from nations that had been unsuccessful in solving their internal economic problems, but the crucial short-run internal challenge now, he noted, was unemployment rather than overpopulation. The causal sequence, nevertheless, was the same. The route to world peace lay in internal macroeconomic reform and in international economic cooperation.

Victory for the Economists in the Second World War

The economics discipline entered the Second World War without having thought deeply over the prior two decades about how to respond to conflict. Certainly it had no sub-discipline, as it does today, concerned with the economics of war, peace, defence and security. Moreover, economics was held in relatively low repute, at least among those in power. It was perceived, correctly, as an eclectic multi-paradigmatic field that spoke still with many voices: marginalist, Institutionalist, Keynesian, socialist, historical and others. Moreover, several of these voices were to a degree threatening to the established order. They came in America from advisers to US President Franklin D. Roosevelt, friends of organized labour and farmers, from architects of a welfare state – and worse. On the American campus in the 1930s academic freedom cases frequently involved economists, the result of the authorities attempting to silence troublemakers (Hofstadter and Metzger 1955). It was ironic that by 1945 the Second World War had given to economics a much improved reputation and a perception among outsiders of a social value that had not been recognized before. Arguably, it gave as much to economics as economics gave to it.

Even though few economists had made a study of war before 1939 they seemed to know what to do when the war began. Undoubtedly the memory of poor administrative performance during the First World War caused governments to welcome economists with relatively open arms 25 years later. In the first war resources and finished goods had been squandered by the bureaucrats and business persons recruited for the occasion. War finance had stumbled, prices had risen unacceptably, and profiteering occurred to a scandalous degree. Now, the economists claimed, they could do it all better.

It turned out that in the Second World War economists of all stripes could find useful work. Moreover, they got along rather well with the public servants and a new set of volunteers from the private sector. Institutionalists like John Kenneth Galbraith supervised price control and war mobilization (Galbraith 1981). Keynesians like Lauchlin Currie and Gerhard Colm designed non-inflationary monetary and fiscal policy (Sandilands 1990; Stein 1969). Microeconomists like Charles Kindleberger (1991) demonstrated the use of optimizing techniques in selecting bombing targets, and economic historians like Walt Rostow helped to create the first high-powered intelligence service (Lodewijks 1991). Even smart young Marxists like Paul Baran and

Paul Sweezy found useful roles. A consequence of the impressive wartime performance by economists was that, when the war ended, the discipline had gained new respect. Indeed, politicians and bureaucrats alike found the economist's perspective so useful that they searched for ways to embed the discipline in government in peacetime. The Council of Economic Advisers, set up in the White House, and the semi-public RAND Corporation were both established with this objective in mind. After the war economists were sprinkled liberally throughout both the executive and legislative branches and the Federal Reserve, and the economist's approach was seen in the consideration of problems both of peace and war. The placement by Secretary of Defense McNamara of 'whiz kids' – economists and 'systems analysts' – firmly by his side when he entered the Kennedy administration in 1960 was only the latest step in the infiltration by economists that had started 20 years before.

When the National Science Foundation was created after the Second World War in response to Vannevar Bush's persuasive picture, *Science, the Endless Frontier* (1945) it was symbolic that the case for inclusion of economics was made not on the basis of its theoretical innovations or contributions over two centuries to human welfare; instead, a list was presented of the contributions to the recent war effort. The sociologist/economist Talcott Parsons, who was charged by the Social Science Research Council to prepare a lobbying document for use with Congress, described the participation of economists in such new agencies as the Office of Strategic Services and the Foreign Economic Administration, as well as the strengthening of older ones such as the Bureau of Agricultural Economics and the Treasury. The most impressive contribution of economists, Parsons concluded, came in the application of the Keynesian theory of income determination, primarily by Simon Kuznets, to achieve wartime full employment without inflation (Klausner and Lidz 1986, p. 100).

The discussion in this entry thus far has been mainly about how economics has addressed war throughout its history, but after 1945 the relationship ran in both directions. Economics was itself changed fundamentally by its involvement with war though contacts with new people, new problems, new professional circumstances and new funding sources such as the Office of Naval Research. The most revolutionary new tool brought into economic analysis during the war was game theory, emerging from the collaboration of the mathematician John Von Neumann with the economist Oscar Morgenstern (Weintraub 1992). Used originally to model parties in conflict it was extended to models of interactions between all kinds of actors, including those in cooperation. Operations research was another field with deep military roots to which economics contributed, and from which it brought back tools to the discipline (Mirowski 2002).

New global breadth was the third contribution of war to economics. Not only were conventional issues of international trade and finance given new urgency by the challenge of rebuilding the world economy, but for both economic and strategic reasons it became necessary to understand friends and foes more thoroughly. This gave rise to new fields called 'comparative economic systems' and 'Soviet studies'. When it became necessary to organize recovery from the Second World War and construct strong nations as bulwarks against Communism, 'development economics' appeared as a new sub-discipline, with generous funding from government and private foundations.

The Triumph of economics in the Cold War

The rise to prominence of economics in the study of war began during the Second World War. The discipline had shown that it could be useful in many ways during full mobilization. But the complete edifice of modern defence economics was constructed only later, during the Cold War (1946–89). There were three reasons. Perhaps of greatest significance was the steadily increasing cost of weapons systems. More items than just nuclear weapons and space exploration bore big price tags – such things as carrier battle groups, strategic bombers and foreign basing of forces. Clearly, hard choices had to be made among alternatives. Yet the questions remained: who should

make the choices and how should they be made? Conflict of interest might become rampant. Decision-making by politicians was not appealing, since they were likely to put the prospect of re-election ahead of the public interest, and maybe even of human survival. Military forces protected the republic, but from the perspective of a legislator they also resembled pork. The military contractors in the private sector were just as problematic: profit rather than national interest usually dominated their incentive structures, even when on leave in government. Despite legislation that established a unified command under the Joint Chiefs of Staff, leaders of the separate forces still engaged in wasteful competition for resources and resisted cost-saving cooperation. When the warrior president Dwight Eisenhower echoed Erasmus and warned of the dangers from a rising 'military–industrial complex' there was an unassailable cause for concern.

So enter the economists. They announced that optimization subject to constraints was their specialty. Scarcity was their stock in trade. Pure reason was their method – and they were incorruptible. They accepted neither ideology nor any sentimental appeals to service loyalty or norms other than the public welfare. They were trained to look for rent seekers under every bed. Voters accepted the necessity of an adequate defence in the Cold War but they wanted as little of it as possible to get the job done, and they wanted it at the lowest cost. Economists made a plausible case that they were best equipped to get the job done that way.

A second reason for the rise to power of economists during the Cold War was their evident command of analytical techniques critical to the management of nuclear confrontation. The skills of the economist in minimizing the costs of defence were certainly comforting, but their avowed capacity to manage weapons of mass destruction was valued even more. The essential consideration was that if nuclear weapons were ever used again this would constitute profound policy failure. Bluff, restraint, and an absolute reluctance to engage simply had to be the main characteristics of nuclear policy. But surely these were not the characteristics that were ingrained in the professional soldier, who was trained to fight – precisely the action that could not be countenanced when the weapons employed would kill friend and foe alike? Indeed, the Cold War was the economist's war par excellence, fought at last with the condition wished for by Adam Smith 200 years before that the powerful nations of the world would reach a stage when conflict of any kind would be 'irrational' and the role of strategy was exclusively to arrange for war never to occur. Deterrence had to be on a massive scale and every step taken to see that there could not be a miscalculation. The first mistake could be fatal. The memory of Sarajevo, and world war by accident, was very vivid still. Economists made a strong case for their role in military strategy and operations because, above all, they seemed to recognize the absolute necessity of peace during the Cold War.

The third reason for the rise to prominence of economists in the Cold War, and for their fingers rather than those of more conventional warriors to be on the trigger, was the alternatives. In the post-war years an image emerged of high-placed military commanders that was deeply troubling; in important cases they seemed arrogant, vain, and sometimes out of control. As events unfolded and memoirs were published, Generals Patton, De Gaulle, MacArthur, Montgomery and others all appeared to be in varying degrees emotional prima donnas, insubordinate to civilian control. These men might be much beloved of their troops, and they might have tactical skills and 'leadership qualities' necessary to triumph in conventional battle, but could they be trusted with the ticket to Armageddon? Many voters said emphatically 'No!' Closer in time, by the 1950s, than the mythic warriors of the Second World War was General Curtis Lemay, chief of the United States Strategic Air Command, who talked of bombing the enemy back to the Stone Age. Worried citizens might be pardoned for wondering whether, if ever he took such action the enemy might not find the means to bring everyone else along with them. The fictional extrapolation of General Lemay, represented in the film *Dr Strangelove,* sent many a shiver down American spines. The Cuban missile crisis in 1962 also provided a

real-life example of why the fate of the world should rest on cool heads and calculating minds. Rightly or wrongly, as early as the 1950s many inside and outside of government had concluded that military affairs had become too dangerous to be left to the military. But who else was there? The political leaders did not seem able to rise to the challenge. The brass might grumble at the prospect of military decision-making falling into the hands of bloodless civilians, but to others this location promised relief. Among the civilians the economists, appreciated now for their dispassion and detachment, seemed the most attractive candidates. In constructing policy toward Cuba the rumpled game theorist and Harvard professor Thomas Schelling (Nobel Prize 2005) seemed a much preferred alternative to Dr Strangelove.

Economists entered the highest reaches of military strategy by several routes. After the National Security Act of 1948 in the United States provided for establishment of a National Security Council, economists began to take important positions on the Council staff. The Harvard economist Carl Kaysen became Deputy National Security Adviser in the Kennedy administration and Walt Rostow from the Massachusetts Institute of Technology (MIT) became National Security Adviser to President Johnson. Robert McNamara, Secretary of Defense under both Kennedy and Johnson, while not a professional economist himself had a high regard for the economist's approach. He set up a systems analysis unit in the Pentagon under the economist Alain Enthoven that was intended to carry the economist's way of thinking throughout the military. The RAND Corporation was an important Cold War resource with advice flowing at the highest levels from such well-known economists as Henry Rowen, James Schlesinger, Stephen Enke and Charles Wolf. The academic discipline of the time also contributed a galaxy of stars who served as consultants and informal advisors to various parts of the legislative and executive branches. Thomas Schelling, Charles Hitch, Roland McKean and Martin Shubik are only a few of the most prominent. None of these considered himself a specialist exclusively on war but rather a general economist who could turn his attention, and his tools, to the particular problems presented by war. The appointments of James Schlesinger as Secretary of Defense and director of the Central Intelligence Agency, both in 1973, signalled the zenith of economics in the US defence establishment.

An intriguing characteristic of the economists involved with the Cold War was their unwillingness to hive off into a regularized subsection of the economics discipline, such as health or agricultural economics. Indeed, they often led a life of glamour and glory unknown to other economists before or since. They operated near the apex of power, were cleared at the highest level of security, and dealt directly with urgent matters of life and death to the nation. Moreover, they were compelled to become multidisciplinary if they were to operate effectively and to become part of the informal community that contained government officials of all kinds, business leaders, scientists, and members of the other disciplines concerned with war, including history, law, philosophy, political science, sociology and others. They had to keep abreast of arcane technological developments and new weapons, learn about treaties, and comprehend the relations of the armed forces with society. They became part of the field known as 'strategic studies' (served by the London-based International Institute of Strategic Studies), students of international relations, and regular contributors to the mass media. They might see more of leaders in other defence-related fields, such as Bernard Brodie, Alistair Buchan, Hedley Bull, Harold Brown, Richard Garwin and Albert Wohlstetter, than they did of their disciplinary kin.

While some economists rose to positions of great influence within the defence and foreign policy establishment during the Cold War, others became involved in what was known sometimes as peace and conflict studies, or peace science. Prominent economists in this movement included Kenneth Boulding and Walter Isard. Their focus was particularly on costs of defence, causes of conflict, arms races, and arms control and disarmament. Their influence seems to have been confined mainly to their own community.

The landmark study in the defence economics literature of the cold war is *The Economics of*

Defense in the Nuclear Age (1960) by C.J. Hitch and Roland McKean. This work codified much that was known already, made original contributions itself and pointed to research opportunities ahead. Two directions into which other economists moved were towards an understanding of arms races, led by the pioneering work of Lewis Richardson *(Arms and Insecurity,* 1960), Walter Isard and others, and towards an appreciation of the impact of defence expenditures on economic growth, led by Emile Benoit's *Defense and Economic Growth in Developing Countries* (1973).

To many in the home discipline of economics the defence economists of the 1950s, 1960s and 1970s must have seemed a strange breed, dancing to a different drummer from those with whom they had studied in graduate school. In fact the circumstances of these defence economists have had few parallels in the history of the discipline, and they did not last for long. Perhaps the debacle in Vietnam and the wind-down of the Cold War explain their loss of influence. Or perhaps they were not well equipped for the challenges of the 1980s when the Reagan administration in the United States and the Thatcher government in Britain talked of defeating the enemy rather than avoiding conflict. Moreover, the warfare that loomed was embedded in ethnic, religious and historical grievances about which economists knew rather little.

Birth of a Sub-discipline

As economists retreated from the pinnacle of power in the 1980s, down from cabinet and sub-cabinet positions to advisors, analysts and consultants, a new sub-discipline of 'defence economics' emerged. In the words of two of its founders it was concerned with such topics as 'the analysis of alliance burden sharing, the effects of contract design on procurement, the impact of defence expenditures on economic growth, and the economic consequences of arms control treaties' (Sandler and Hartley 1995, p. 1). This new applied sub-discipline was comparable to the older agricultural and labour economics, and to such newer ones as health and environmental economics. Defence economists adhered to the core theory accepted in the mother discipline, and they applied approved analytical tools to problems that were perceived to be susceptible to the economists' methods. What distinguished these applied fields was knowledge of institutional facts in particular areas and some attention to policy issues prominent there. Sandler and Hartley, authors of a survey of the literature in the mid-1990s, described the field as follows: 'Defence economics involves the application of economic reasoning and methods to the study of defence- related issues. Defence economics differs from other fields of economics in at least three ways: (1) the set of agents studied (e.g. defence contractors, branches of the military); (2) the institutional arrangements of the defence establishment (e.g. procurement procedures); and (3) the set of issues investigated' (Sandler and Hartley 1995, p. xi). They suggest that economists explore four 'basic economic problems': allocative efficiency, public choice considerations (why elected officials and bureaucrats behave the way they do), the distributive implications of defence decisions, and stability issues, including paths after shocks. A specialized journal, *Defense and Peace Economics,* was founded in 1990 to weld together two strands that during the cold war had gone in rather different directions.

Sandler and Hartley recognize that recent events have shifted attention away from long-standing concentration on superpower confrontation to a variety of new issues such as the consequences of the break-up of the socialist world, the increased number of regional conflicts such as the Gulf War of 1991 and subsequent Iraq War that followed the relaxation of superpower hegemony, the proliferation of relatively inexpensive conventional weapons (some second-hand), the responsibilities of the arms exporters, the design of arms control treaties, and the macroeconomic implications of military downsizing. In the 1990s there was even talk of a peace dividend – an idea that seems merely poignant in the 21st century. Terrorism, guerilla warfare, the security implications of economic coalitions, and burden sharing within alliances remained topics on which economists

found they had interesting things to say. Some issues, like procurement practices, are common to all parts of the public sector but receive disproportionate attention in defence economics because of the magnitude of defence expenditures.

A textbook for an undergraduate course on 'The Economics of War' published in 2006 (Poast 2006) lists aspects of international conflict that may be understood better using the economist's tools: how to achieve mobilization or disarmament; the special problems presented by recruitment of a military labour force and weapons procurement; conflict in developing countries, the small-arms trade, peacekeeping; and the dilemmas of terrorism and the proliferation of the weapons of mass destruction in the 21st century.

John Maynard Keynes remarked in 1930 that he looked forward to the day when economists would be 'thought of as humble, competent people, on a level with dentists' (Keynes 1930, p. 332). It appears that in the 21st century economists concerned with the study of war have gained this status. It is now well accepted that war, like all human activity, requires the recognition of scarcity and the need to make choices based on forgone opportunities. This is the domain of economics. Defence economists stand ready to advise on these allocative decisions and to remind policymakers of the applicability of such concepts as externalities and public goods. The heroic years of defence economics are almost certainly gone for ever; the economists are today, as they say, on tap but not on top. Nevertheless, their usefulness remains, even if at a more modest level than before. The study of war is now an accepted part of economics, assigned to its own sub-field and dependent heavily on the tools and methods of public economics. In its current posture economics is less likely to find a cure for conflict than to make it more efficient and its prevention less costly. In a world full of shortages and sufferings this is no small accomplishment.

See Also

► Defence Economics

Bibliography

Angell, N. 1910. *The great illusion: A study of the relation of military power in nations to their economic and social advantage*, 3rd ed, 1912. New York/London: G.P. Purnam's Sons.

Barber, W. 1991a. British and American economists and attempts to comprehend the nature of war, 1919–20. In Goodwin (1991).

Barber, W. 1991b. From the economics of welfare to the economics of warfare (and back) in the thought of A.-C. Pigou. In Goodwin (1991).

Benoit, E. 1973. *Defense and economic growth in developing countries*. Boston: D.C. Heath.

Biddle, J.E., and W.J. Samuels. 1991. Thorstein Veblen on war, peace, and national security. In Goodwin (1991).

Black, R.D.C., A.W. Coats, and C.D. Goodwin. 1973. *The marginal revolution in economics: Interpretation and evaluation*. Durham: Duke University Press.

Bush, V. 1945. *Science, the endless frontier*. Washington: US Government Printing Office.

Clark, J.M. 1931. *The costs of the world war to the American people*. New Haven: Yale University Press.

Davis, C.M. 1991. Marxist and Soviet defense economics, 1848–1927. In Goodwin (1991).

De Marchi, N. 1991. League of Nations economists and the ideal of peaceful change in the decade of the thirties. In Goodwin (1991).

Galbraith, J.K. 1981. *A life in our times: Memoirs*. Boston: Houghton Mifflin.

Goodwin, C.D. (ed.). 1991. *Economics and national security: A history of their interaction*. Durham: Duke University Press.

Hitch, C.J., and R.N. McKean. 1960. *The economics of defense in the nuclear age*. Cambridge, MA: Harvard University Press.

Hofstadter, R., and W.P. Metzger. 1955. *The development of academic freedom in the United States*. New York: Columbia University Press.

Keynes, J.M. 1920. *The economic consequences of the peace*. New York: Harcourt Brace.

Keynes, J.M. 1930. Economic possibilities for our grandchildren. In *The collected writings of John Maynard Keynes*, ed. D. Moggridge. New York: Macmillan, 1972.

Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan.

Kindleberger, C.P. 1991. *The life of an economist: An autobiography*. Cambridge, MA: Basil Blackwell.

Klausner, S.Z., and V.M. Lidz. 1986. *The nationalization of the social sciences*. Philadelphia: University of Pennsylvania Press.

Leonard, R. 1991. War as a 'simple economic problem': The rise of an economics of defense. In Goodwin (1991).

Lodewijks, J. 1991. Developing economies, and national security policy. In Goodwin (1991).

Lowry, S.T. 1991. Preclassical perceptions of economy and security. In Goodwin (1991).

Malthus, T.R. 1826. *An essay on the principle of population*, 6th ed. Vols. 2 and 3 of *The works of Thomas Robert Malthus*, ed. E.A. Wrigley and D. Souden. London: Pickering and Chatto, 1986.
Mill, J.S. 1848. *Principles of political economy.* Vols 2 and 3 of *The collected works of John Stuart Mill*, ed. J.-M. Robson. Toronto: University of Toronto Press, 1965.
Mirowski, P. 2002. *Machine dreams: Economics becomes a cyborg science*. New York: Cambridge University Press.
Pigou, A.C. 1921. *The political economy of war*. London: Macmillan.
Poast, P. 2006. *The economics of war*. New York: McGraw Hill.
Rae, J. 1834. In *Statement of some new principles on the subject of political economy*, ed. R. Warren James. Toronto: University of Toronto Press, 1965.
Richardson, L. 1960. *Arms and insecurity.* Pittsburgh: Boxwood Press.
Sandilands, R. 1990. *The life and political economy of Lauchlin Currie*. Durham: Duke University Press.
Sandler, T., and K. Hartley. 1995. *The economics of defense*. Cambridge: Cambridge University Press.
Silberner, E. 1946. *The problem of war in nineteenth century economic thought*. Princeton: Princeton University Press.
Smith, A. 1776. In *An inquiry into the nature and causes of the eealth of nations,* ed. R.H. Campbell and A.S. Skinner. Oxford: Clarendon Press, 1976.
Speelman, P.J. 2002. *Henry Lloyd and the military enlightenment of eighteenth- century Europe*. Westport: Greenwood Press.
Stein, H. 1969. *The fiscal revolution in America*. Chicago: University of Chicago Press.
Weintraub, E.R. 1992. *Toward a history of game theory.* Durham: Duke University Press.

War Economy

Avner Offer

In wartime, many markets are suspended. When survival is at stake, governments arrogate the task of setting priorities and the power to allocate resources. The problems of wartime allocation are typically those of management and politics, and have to be resolved ad hoc. The majority of wars since 1945 have taken place in less-developed economies, often in the form of insurgencies or civil wars; but this experience and its literature are both fragmented. What follows is based largely on the experience of Germany, Britain and the United States in the First and Second World Wars, in Korea and in Vietnam.

Wartime priorities, even the choice of war as a policy, reflect the constraints of the economy's endowments. At the pre-war stage of power-building, leaders assess their prospective opponents and choose the most appropriate forms of power: whether it should be intensive in capital, labour, enterprise or land. Economic choices are made between bayonets, warships and aircraft, between numbers and quality, home procurement and imports, firepower and mobility, regular war and insurgency, paid armies, voluntary forces or conscription, human capital versus technology and material, defensive versus offensive, fortification versus manoeuvre, 'total war' versus 'limited liability'. Productivity varies in destruction as well as production, and combatants use similar resources with different results. Enterprise, skill and motivation are even more decisive than they are in peacetime. If strategies are influenced by factor endowment, the choice of strategy, in its turn, determines the nature of the economic problem.

One strategic option that has been attractive in the 20th century is economic warfare. This consists of an attempt to identify and destroy vital links in the enemy's economy. World War I the Allied blockade of Germany helped to bring about the collapse of the food supply system, while Britain narrowly avoided the same fate in both world wars. Britain and the United States (but not Germany and Russia) developed doctrines of strategic bombing and dedicated bomber forces to attack the German and Japanese economies. American bomber forces concentrated on economic bottlenecks, and shifted their attack in sequence from key industrial plants to the oil distribution and the transport systems. Britain's Bomber Command, operating at night with less precision, took whole cities as its targets, in the hope of destroying the housing stock and civilian motivation. The flexibility and range of substitutions available in a economy meant that none of these efforts was decisive on its own, and the

contribution of economic warfare to Germany's collapse is still debated.

In a war economy the state typically controls a much higher proportion of the national product than in peacetime. One half or more of output can be diverted to military uses. Business rarely continues as usual: instead, government acquires legal powers to direct production, distribution and exchange. But its ability to do so effectively only develops gradually. By taking up idle resources war management can raise the level of output considerably; civilian consumption, especially of durables, is also easy to sacrifice. The existing economic shell remains in place, to make use of production and managerial skills. It is important to reduce duplication of effort, but competition helps to maintain innovation, so firms receive large and secure contracts but retain some independence. War production was largely amenable to the methods of mass production, and often exceeded the scale achieved (or indeed conceivable) in civilian market production. Cost was a secondary consideration. Working at full capacity to produce standard articles at guaranteed prices allowed manufacturers to retain large profits and invest in new plant. Where the difficulties of production exceeded the capacity of private industry (in the case of explosives in the UK, and atomic weapons in America) the government set up or expanded its own arsenals. Much of the new capacity was not easily converted to market production and remained a burden on postwar balance sheets.

Large corporations in America had already undertaken the central management of diversified enterprises before World War I. These methods were adapted to industry-wide management boards in all three countries. Businessmen came forward to manage whole sectors of the economy, often affirming their primary allegiance by refusing to take more than 'a dollar a year'. Such boards commonly governed transport, mining, metals, shipping and food and raw material distribution. The methods used combined delivery quotas, physical resource allocation, price fixing and market incentives. This symbiosis of business and government was known, ironically, as 'war socialism'.

Military tactics in the two world wars were manpower-intensive, and labour was a more crucial factor than capital. The choice between productive and military employment was stark. On the continent of Europe conscription was established in the 19th century and still remains the rule. Britain entered the first war with a professional army, went over to volunteers, and finally, like its allies overseas (except Australia) introduced conscription. All manpower systems gave preference to military over civilian requirements, often to the detriment of efficiency: skilled technicians sometimes served as riflemen. But much of the labour taken up by the armies could be replaced by women, and mass production lent itself to the dilution of skills. In both world wars manpower became a serious constraint on the combatants, especially Germany, which ran short in the trenches, the factories and the farms; in World War II Germany pressed in millions of forced labourers from prison camps and occupied territories. Full employment in large-scale industry is conductive to trade union organization, and memberships rose to record levels. For their part, unions had to remove restrictive practices and repudiate strikes. In return, the membership was shielded from industrial and often military conscription, while the leadership got recognition, incorporation and even partnership in government. Such accommodations weakened the hold of unions over their members, and when wages and working conditions failed to keep up with prices, unauthorized stoppages and unrest kept recurring; in World War I, unrest was transformed into revolution in the wake of military defeat.

War increases the consumption of foods, fibres, other raw materials and metals. It diverts resources from manufacturing and agriculture and creates a general state of material shortages. Agriculture presents some of the most difficult problems of management. Its manpower and draft animals were mobilized by the military, and in Europe it suffered from shortages of imported fodder and fertilizer and breakdowns in transport. In Germany, the official reaction to food shortages was to impose maximum consumer prices and delivery quotas, which helped to reduce shipments to well below peacetime levels, as

producers chose to consume more of the output themselves. In contrast, in English-speaking countries farmers got minimum prices, delivered record harvests and achieved a good balance between livestock and arable products. All combatants adopted rationing. In the name of solidarity and fairness, rationing systems tended to equality; often as in the case of Germany, to the point of ignoring physiological inequalities. Rationing, not only of food but of most other consumer commodities, has called for major efforts of administration, which were sometimes tainted by corruption. Extensive black markets restored differentials by offering restricted goods illegally at high prices. With housebuilding at a standstill, all combatants kept inflation and unrest down by holding dwelling rents below market prices. These controls often persisted years beyond the end of the war, distorted pre-existing housing supply systems and affected the structure of the industry.

International trade assumes a peculiar quality in wartime. Transactions become one-sided and unequal, and commodities trade for political assets and capital transfers. In both world wars Britain ran up very large deficits overseas and also liquidated many of its overseas investments. In its turn, Britain supported its allies extensively, while American economic aid underpinned British survival in both world wars. In the First World War, and more systematically in the Second, Germany plundered its occupied territories. The large imbalances of international trade broke down the convertibility of currencies, and placed obstacles on the road to recovery.

Transport is a weak link. Railways are rigid and not easily adapted to wartime freight flows, while shipping is a very lumpy form of capital, which was destroyed and reproduced wholesale in the two world wars. In the second, railways were also disrupted from the air. Transport equipment competed with munitions for the same scarce labour, machinery and materials. In consequence, transport bottlenecks hampered both the military and civilian war efforts. Like other consumer durables, the manufacture of motor cars almost ceased and petrol was strictly rationed, thereby imposing further handicaps on the transport system.

Arms races are a permanent fixture of the 20th century, and the competition of weapons accelerates in wartime. This stimulates technical innovation, and war is a fertile source of practical and impractical inventions. Shortages have prompted technological substitutions like air-fixed nitrogen and synthetic foodstuffs, and social ones like daylight saving. War also effected technology transfers as patents were suspended for the duration, and often for several years beyond.

Finance is a key problem of economic management. The origins of central banking and national debts lie in the war finance of the 18th century. The problem is simply stated, if not so simply resolved. The state withdraws labour and commodities from the economy, and pays with liquid funds. An increased supply of money raises the prices of a depleted flow of civilian goods and services. To keep inflation in check, the state has to withdraw liquidity as fast as it pumps it in. Taxation rises, but its effectiveness depends on civilian motivation, and war governments have preferred to raise most of their funds by borrowing; some of it absorbs purchasing power from individuals and firms, while borrowing from the central and commercial banks increases liquidity. Requisition and confiscation are less efficient and are only used in exceptional circumstances. In both world wars direct taxes rose by an order of magnitude, while war loans paid rates of interest well above prewar levels. Direct taxation penetrated down the social scale, and one of the best British innovations of World War II was deferred wages. As wars progressed, businessmen were made to pay special taxes on their superprofits.

None of these measures could square the circle. In World War I prices more than doubled in four years. Inflation ran completely out of control in the defeated countries and was only arrested in allied countries by savage deflation in 1920–21. The lesson was learned, and in World War II inflation was largely kept in check by measures of forced saving. National debts increased by an order of magnitude. In World War I, the combatants also pinned some hopes on a large indemnity, for which the Franco–Prussian war of 1870–71 provided a precedent. But reparations proved to

be a mirage, and neither the allies, nor the Americans who supported them, managed to effect a sufficiently large transfer from debtors to creditors in the inter-war years. After World War II the effort was not repeated; the United States pumped loans and grants into Western Europe during the reconstruction period. This allowed a very rapid recovery in Germany and Japan, which was also assisted by the destruction of their obsolete industrial plant.

20th-century wars have had considerable distributional effects, both international and domestic. From an economic point of view both world wars have favoured the United States, allowing it to capture new markets, wiping out debt in the First World War and pulling the economy out of depression in the Second. Two subsequent American wars, in Korea and Vietnam, were not so beneficial: with the economy already close to full capacity, they fuelled inflation and may have crowded out domestic projects. Domestically, wartime favours the productive sectors: Industrialists, skilled factory workers and farmers found themselves in positions of relative strength. Labour benefited: after a short period of frictional unemployment at the outset, full employment ensued; even if wages failed to keep up with prices, family earnings usually ran well ahead, with longer hours and higher participation rates, male and especially female. Rationing did not prevent a substantial improvement in working-class nutrition in Britain, while in Germany, for example, a slight relative improvement was wiped out by the absolute immiseration of civilian society. The inter-war settlement in Britain effected a redistribution to capital as women withdrew from the workforce, unemployment reappeared and a large share of tax revenue was transferred to the owners of the national debt. In Germany and central Europe inflation wiped out financial assets and the national debt altogether, and impoverished the owners of fixed-interest securities. It is often argued that wartime laid the ground for a more generous welfare policy post-war, in a number of ways. It began in the form of transfers to servicemen's families and continued as veterans' benefits. More debatable is the claim that mass-participation wars give rise to expectations of reward. 'Homes fit for Heroes', the World War I slogan, captures this expectation; a commitment to full employment followed World War II in both Britain and America. But government expenditure post-war merely takes up the prewar trend, albeit at higher levels. Certainly wars work to raise acceptable tax levels and involve the State in extensive areas of social policy.

War economies build up before the shooting begins and they continue after it subsides. After World War I government share of GNP fell in most combatant economies, although taxation remained high. Government expenditure picked up during the rearmament phase of the 1930s in Germany and Britain. After the end of World War II America and Russia remained in a state of semi-mobilization. One feature is a high share of defence expenditure as a share of GNP; in the United States this is about seven times the inter-war level. Another is conscription, which remains universal in continental Europe, and was only dismantled in the United States after Vietnam. It is argued that defence technology 'crowds out' civilian enterprise, as a good part of scientific manpower works in military research. High defence expenditures continue to exert inflationary pressure. Another, more debatable aspect, is the continuing welfare orientation of advanced economies, which perhaps owes something to wartime solidarity, tax levels and governmental powers. Business symbiosis with defence continues. In America the heads of both General Motors and Ford became Secretaries of Defense, and some large corporations depend heavily on military budgets. One persistent feature of wartime economy in peace are stockpiles of strategic commodities, which are sometimes used to regulate production and prices. Farm support programmes hark back to the experience and apprehensions of the war economy. Limitations on technology transfer are justified not merely by commercial expediency, but by national security.

The command economy of Soviet Russia was formed under the shadow of military threat and consequently has many features of a war economy: a very large military sector, conscription, shortage of consumer goods and housing coupled with strictly controlled prices, high female

participation rates, a tendency to autarky, extensive (but low-quality) welfare services and low unemployment. On the other hand it has successfully controlled inflation by means of price and wage controls, and by retaining financial assets almost entirely in state hands.

War is common enough as a policy choice, but is really too uncertain to be approached entirely as a rational undertaking. The stakes and imponderables are very large. Both sides (if there are only two) assume in advance that fighting is worthwhile and at least one of them is wrong. Economic theory has consequently not found much application in war, although the use of price incentives by American and British war administrators have, on the whole, proved sounder than the physical and administrative controls of the Germans. World War II saw the attempts to optimize military decisions by means of operations research and empirical survey. An attempt to manage war with economic tools and systematic analysis of costs, risks and benefits was started at the American Department of Defense under Robert McNamara in the 1960s. Suffice it to say that the best available theory and the world's most powerful economy were not equal to the task.

See Also

- ▶ Defence Economics
- ▶ Military Expenditure
- ▶ Price Control

Bibliography

Carnegie Endowment for International Peace. *Economic and Social History of the War*. (Many vols, covering most of the European combatants and some neutrals in World War I. This is the largest body of published research on war economies.)

Chester, D.N. (ed.). 1951. *Lessons of the British war economy*. Cambridge: Cambridge University Press.

Clarkson, G.B. 1923. *Industrial America in the World War: The strategy behind the line 1917–1918*. New York: Houghton Mifflin.

Enthoven, A.C., and K.W. Smith. *How much is enough? Shaping the defense program, 1961–1969*. New York: Harper & Row.

Hancock, W.K., and M.M. Gowing. 1949. *The British war economy*. London: HMSO. (See also other volumes in the British Official History of the Second World War, Civil Series.).

Hardach, G. 1978. *The first world war 1914–1918*. London: Allen Lane.

Hoffman, F.S. 1959. The economic analysis of defense: Choice without markets. *American Economic Review* 49(2): 368–376.

Kennedy, G. 1983. *Defense economics*. New York: St Martin's Press.

Keynes, J.M. 1940. *How to pay for the war*. London: Macmillan.

Klein, G.H. 1959. *Germany's economic preparations for war*. Cambridge, MA: Harvard University Press.

Knorr, K. 1956. *The war potential of nations*. Princeton: Princeton University Press.

Koistinen, P.A.C. 1980. *The military-industrial complex: A historical perspective*. New York: Praeger.

Lloyd, E.H.M. 1924. *Experiments in state control at the War Office and the Ministry of Food*, Carnegie Endowment for International Peace, Economic and Social History of the War, British series. New Haven: Yale University Press.

Milward, A. 1967. *The German economy at war*. London: Athlone Press.

Milward, A. 1976. *War, economy and society 1939–1945*. London: Allen Lane.

Olson Jr., M. 1963. *The economics of wartime shortage. A history of British food supplies in the Napoleonic War and in World Wars I and II*. Durham: Duke University Press.

Scitovsky, T., E. Shaw, and L. Tarshis. 1951. *Mobilizing resources for War: The economic alternatives*. New York: McGraw-Hill.

Skalweit, A. 1927. *Die Deutsche Kriegsernahrungswirtschaft*, Carnegie Endowment for International Peace, Economic and Social History of the War, German series. Stuttgart: Deutsche Verlagsanstalt.

United States Strategic Bombing Survey. *The effects of strategic bombing on the German War economy*. Washington: GPO.

Warburton, Clark (1896–1979)

L. Yeager

Keywords

Monetarism; Monetary disequilibrium; Quantity theory of money; Real business cycles; Warburton, C.

JEL Classifications
B31

American economist; pioneer, before Milton Friedman, in research later labelled 'monetarist', and a critic of Keynesianism during the years when that doctrine was crowding out attention to money.

Warburton was born on 27 January 1896 near Buffalo, New York, and died on 18 September 1979 in Fairfax, Virginia. After overseas military service during the First World War, he earned bachelor's and master's degrees from Cornell University. He published his 1932 Columbia Ph. D. dissertation as *The Economic Results of Prohibition.* He held teaching positions in India and the United States in the 1920s and early 1930s and worked at the Brookings Institution from 1932 to 1934, coauthoring *America's Capacity to Consume.* He then joined the newly organized Federal Deposit Insurance Corporation. Although routine FDIC work consumed much of his time (as his files reveal), he still managed to publish over 30 papers on monetary economics, most of them empirically oriented, from 1943 to 1953. Altered FDIC policy then impeded his research and publication until about 1962, when he took a brief leave to serve with the Banking and Currency Committee of the US House of Representatives. He was elected President of the Southern Economic Association for 1963–4. After retiring from the FDIC in 1965, he taught briefly at the University of California, Davis.

Warburton originally accepted a 'real' theory of the business cycle, but scrutiny of statistical and qualitative history changed his views. Using quarterly as well as annual data, he found that deviations from trend of the quantity of money generally preceded turning points in business conditions (and velocity deviations followed). While accepting a quantity-of-money theory of the price level in the long run, he recognized how elements of wage and price stickiness cause monetary disturbances to impinge on output first; he espoused a 'monetary disequilibrium theory' (which, despite its venerability, has ironically been mislabelled 'Keynesian' in recent years). He understood that disequilibrium does not necessarily imply irrational behaviour by individuals.

Warburton emphasized the role of money and inadequate monetary policy in the Great Depression of the 1930s. He continued to criticize the Federal Reserve for deficiencies in its economic theory and research and, in particular, for relying on interest rates in deciding and implementing policy. He believed that pure fiscal policy, unsupported by changes in the quantity of money, is ineffective as a tool of demand management. Sceptical of the authorities' ability to fine-tune the economy, he recommended a policy of steady growth in the quantity of money at a moderate rate appropriate to trends in the labour force, productivity and velocity.

For Warburton, monetarism was an interpretation reached inductively, not a comprehensive ideology. (So far as any ideology came across in conversations, it was a rather conventional New Deal reformism or liberalism with humanitarian underpinnings.)

Nineteen of Warburton's papers dating from 1945 to 1953 are reprinted, along with a new introduction, in *Depression, Inflation, and Monetary Policy* (1966). Up to his death, Warburton pursued research not only in substantive economics but also in the history of monetary doctrines. These continuing interests are manifest in his last article (published posthumously in *History of Political Economy,* 1981) and in voluminous manuscripts now deposited in the library of George Mason University, Fairfax, Virginia. Plans exist for editing and publishing much of this material.

See Also

► Quantity Theory of Money

Selected Works

1934. (With M. Leven and G. Moulton.) *America's capacity to consume.* Washington, DC: Brookings Institution.

1966. *Depression, inflation, and monetary policy*: *Selected papers, 1945–1953*. Baltimore: Johns Hopkins Press.
1981. Monetary disequilibrium theory in the first half of the twentieth century. *History of Political Economy* 13: 285–299.

Ward, Barbara (1914–81)

Richard Jolly

Barbara Ward was born on 23 May 1914 and died in Sussex on 31 May 1981. After graduating from Somerville College, Oxford, she moved rapidly from teaching and research to journalism, becoming assistant editor of *The Economist* in 1940 and later foreign editor. Although she later held appointments at Harvard and Columbia (from 1957 to 1973) and received numerous honorary degrees, the considerable influence she exercised over four decades on international development thinking and policy was primarily due to her masterful skills in popular communication as journalist, broadcaster, outstanding public speaker and author of a score of best-selling books.

Her contributions to development literature were built around three successive and evolving themes. In *The Rich Nations and the Poor Nations* (1961), she dramatized the wide economic and social gap between the industrial and 'underdeveloped' countries, and underlined the urgent need for international action to bridge it, action for which she argued the Western interest was no less than the interests of the poorer countries themselves. *The Widening Gap* (1971) was a critique of the Pearson Commission report *Partners in Development* (1969), arguing that not only aid but more fundamental changes in international trade, financial arrangements and other economic relationships were also needed. In this respect, *The Widening Gap* laid some of the intellectual foundations for the North–South dialogue of the 1970s and of the subsequent Brandt Commission Report, *North–South: a Programme for Survival* (1980).

Barbara Ward's second contribution is well captured by her concept of 'spaceship earth', her evolving preoccupation with the physical unity and fragility of the planet and what this requires of national and international policy. She emphasized the need for a global perspective of 'human ecology', conservation, the risks of rising armaments and the broader issues of development strategy. These themes were most coherently developed in *Only One Earth: the Care and Maintenance of a Small Planet* (1972) an unofficial report prepared for the United Nations Stockholm Conference on the Human Environment. They also underlay the *Home of Man* (1975) and *Progress for a Small Planet* (1979). Indeed, four or five of the major United Nations' conferences on global issues of the 1970s owed an important part of their intellectual vision and vitality to Barbara Ward.

The third and most persistent characteristic of her contribution to development thinking was the need for vision and the optimistic conviction that enlightened action was almost always possible and could be made politically realistic. The Marshall Plan of 1948–52, was to her a supreme example of such vision and enlightened leadership. Under this plan, the United Stated had for four years transferred some 2 ½% of its GNP on grant terms for the postwar reconstruction of Western Europe. Barbara Ward had eloquently praised this at the time in *The Economist* and she returned to a parallel message in the 1980s when she called for a '20 year Marshall Plan for the Third World'. The call for enlightened leadership, a commitment to morality and idealism in international policy pervaded all Barbara Ward's writings, giving them a cutting edge of practical appeal and inspiration which combined economic liberalism with political radicalism.

Selected Works

1962. *The rich nations and the poor nations*. New York: Norton.

1971. (With J.D. Runnalls and L. D'Anjou, eds) *The widening gap: Development in the 1970's*. New York: Columbia University Press.
1972. (With R. Dubos.) Only one earth: The care and maintenance of a small planet. New York: Norton.
1976. *The home of man*. New York: Norton.
1979. *Progress for a small planet*. New York: Norton.

References

Brandt, W. 1980. *North–South: Program for survival. Report of the Independent Commission on International Development Issues*. Cambridge, MA: MIT Press.
Pearson, L. 1969. *Partners in development: Report of the Commission on International Development*. New York: Praeger.

Warming, Jens (1873–1939)

P. Nørregaard Rasmussen

Keywords
Multiplier; Warming, J.

JEL Classifications
B31

Jens Warming was born on 9 December 1873 and died on 8 September 1939.

After graduating in law, Warming took up economics. In 1919 he became Professor in Applied Statistics at the University of Copenhagen, following more informal attachments to the university. Along with his teaching he produced a number of books describing empirically a wide variety of aspects of the Danish economy. In a way he created the field 'applied statistics' as an academic discipline in Denmark. He not only presented figures, but he surrounded them with reasoning, sometimes naive and not very well articulated, often full of wisdom. One example is his warning of the danger of overfishing because no rent is collected (Warming 1931b). Another most important example is his discovery of the multiplier process, which he presented as early as 1928 and again in 1929–30 and 1931. These important contributions in economic theory were quite often formulated in a somewhat odd way, and they certainly did not attract his fellow economists in Scandinavia.

Warming's formulation of the multiplier runs as follows: assume a closed economy (an assumption he later modified) and consider an investment of, say, 100 units in a railway. This creates an income of equal size, part of it appearing as an increase in savings, but another part as consumption, that latter creating new incomes. This process, he argues, will go on until voluntary savings will increase, so that the newly-constructed railway 'gets an owner' (1929–30). This clearly means that the total voluntary savings in the end equal the impulse, that is, 100 units. An investment will 'finance itself', as he argued at length. However, it does not seem as if Warming was considering the multiplier as part of a more general theory of employment.

Selected Works

1924. *Valutaspørgsmålet* [The problem of foreign exchange].
1928. Beskæftigelsesproblemet [The employment problem]. *Gads Danske Magasin* 22: 609–721.
1929–30. *Danmarks Erhvervs- og Samfundsliv* [Denmark's economic and social life]. Copenhagen.
1931a. Tilpasning [Adaptation]. *Gads Danske Magasin* 25: 481–505.
1931b. Aalegaardsretten. *Nationalkonomisk Tidsskrift* 69: 151–62.
1932. International difficulties arising out of financing of public works during depression. *Economic Journal* 42: 211–24.
1939. *Danmarks Erhvervs- og Samfundsliv* [Denmark's economic and social life]. 2nd ed. Copenhagen: G.E.C. Gad. (This edition was never completed.)

Bibliography

Boserup, M. 1969. A note on the prehistory of the Kahn multiplier. *Economic Journal* 79: 667–669.
Gelting, J.H. 1964. *Jens warming. Nationalkonomisk Tidsskrift*.
Hegeland, H. 1954. *The multiplier theory*. Lund: Gleerup.
Kahn, R.F. 1931. The relation of home investment to unemployment. *Economic Journal* 41: 173–198.

Washington Consensus

John Williamson

Abstract

The original purpose of the term 'Washington Consensus' was to distinguish between a number of policies that remain the subject of partisan controversy and a group of policies that were thought to be consensual in the post-1989 world. After its creation in 1989 the term acquired at least two more meanings. Some used it to describe the policies of the IMF and World Bank, which came to embrace not only institutional reform and a concern with governance but also the two-corners solution for exchange rates and capital account convertibility. Others used the term as a synonym for laissez-faire.

Keywords

Balance of payments; Capital account liberalization; Consensual policies; Corruption; Demographic transition; Deregulation; Development economics; East Asia; Financial liberalization; Fiscal discipline; Foreign direct investment; Globalization; Governance; Import-substituting industrialization; Industrial policy; Inflation; Informal economy; Institutions; Intermediate exchange-rate regime; International Monetary Fund; Laissez- faire; Mont Pèlerin Society; Neoliberalism; Partisan policies; Privatization; Property rights; Public expenditure; Redistribution of income; Tax reform; Trade liberalization; Two-corners solution; Washington Consensus; World Bank

JEL Classification

F

The term 'Washington Consensus' was coined in 1989. The first written usage was in my background paper (Williamson 1990) for a conference that the Institute for International Economics convened in order to examine the extent to which the old ideas of development economics that had governed Latin American economic policy since the 1950s were being swept aside by the set of ideas that had long been accepted as appropriate within the Organisation for Economic Co-operation and Development (OECD). In order to try to ensure that the country papers for that conference dealt with a common set of issues, I made a list of ten policies that I thought more or less everyone in Washington would agree were needed more or less everywhere in Latin America, and labelled this the 'Washington Consensus'. Little did it occur to me that 17 years later I would be asked to write about a term that had become the centre of fierce ideological controversy. On the contrary, I thought of the Washington Consensus as distinguishing between (*a*) what had become consensual and (*b*) what was likely to remain partisan, issues such as income distribution, capital account convertibility, the usefulness of incomes policy, the need to eliminate indexation, the size of the public sector, and the priority to be given to population control and environmental preservation.

The set of 'consensual' policies was:

1. *Fiscal discipline.* This was in the context of a region where almost all the countries had run large deficits, which had led to balance of payments crises and high inflation that hit mainly the poor because the rich could park their money abroad.
2. *Reordering public expenditure priorities.* This suggested switching expenditure in a pro-poor way, from things like indiscriminate subsidies to basic health and education.

3. *Tax reform*: constructing a tax system that would combine a broad tax base with moderate marginal tax rates. This need not imply redistribution of income to the rich if the broadening of the tax base focuses on eliminating loopholes that are exploited by those who can afford to employ tax lawyers.
4. *Liberalizing interest rates.* This was subsequently formulated in a broader way as financial liberalization, to cover also policies like bank privatization and allowing financial institutions to determine the allocation of credit.
5. *A competitive exchange rate.* It was asserted (though this may not have been accurate reporting of the Washington scene) that there was a consensus in favour of ensuring that the exchange rate would be competitive.
6. *Trade liberalization.* It was acknowledged that there was a difference of view about how fast trade should be liberalized, but it was asserted to be widely held that trade needed to be liberalized and countries stood to gain by outward- oriented policies.
7. *Liberalization of inward foreign direct investment.* This specifically did not include comprehensive capital account liberalization, which did not command a consensus in Washington.
8. *Privatization.* This was the one area in which what had originated as a neoliberal (Thatcherite) idea won broad acceptance. We have since been made very conscious that it matters a great deal how privatization is undertaken: it can be a highly corrupt process that transfers assets to a privileged elite for a fraction of their true value, but the evidence that it ultimately brings net benefits is quite strong.
9. *Deregulation.* This focused specifically on easing barriers to entry and exit, not on abolishing regulations designed for safety or environmental reasons.
10. *Property rights.* This was primarily about providing the informal sector with the ability to gain property rights at an acceptable cost.

The term 'Washington Consensus' proved controversial right from the start. Both reformers and critics took it to imply a belief that policy reforms had been imposed by Washington institutions like the International Monetary Fund (IMF) and the World Bank. The reformers resented the implication that they had been following orders rather than implementing policies that they had concluded were needed in their countries. The critics took it as confirmation of their darkest fears. It might have been better if the revolution in economic policy for development had been called something else, but the fact that there was such a revolution is clear. In the 1960s mainstream development thinking held that inflation was at worst harmless and at best a method of extracting resources for investment. Orthodox thought advocated import-substituting industrialization rather than export expansion through globalization. Governments sought to expand the industrial sector by creating more state firms rather than by creating a market-friendly environment. All this changed in the late 1980s, perhaps aided by the collapse of the Communist alternative model, which was widely welcomed in Washington. That was what the Washington Consensus was about.

As time progressed the term came to mean at least three different things. Some people stuck to the original concept. Others used it to mean the evolving policies of the IMF and the World Bank, presumably on the ground that originally it had pretty much described the policies they advocated in 1989. Policy evolved importantly in at least four dimensions. First, in common with development thinking in general, Washington came to understand the importance of institutions in promoting development. Thus, first Naím (1994) and subsequently Burki and Perry (1998) and Kuczynski and Williamson (2003) argued that one needed to supplement the prescriptions of the Washington Consensus with institutional development. Second, the Washington institutions came to place a great emphasis on governance, especially on avoiding corruption. Third, in the IMF in particular opinion moved strongly away from the concern with a competitive exchange rate and the implication of the desirability of an intermediate regime toward the 'two-corners solution' (which argues that to avoid

speculation one should have either a firmly fixed or a freely floating exchange rate, but nothing in between). Fourth, again especially in the IMF, there emerged a strong preference for rapid capital account liberalization. Some of those (like the author) who had been very attached to the original concept of the Washington Consensus found these last two doctrines alien.

The third concept of the term involved an even more radical change from the original, though this is the version that appears to have been widely used by critics. It refers to a programme often described as 'neoliberal' (a term that was at one time used to describe the agenda of the Mont Pèlerin Society, of which Milton Friedman and Friedrich von Hayek were prominent members). In addition to the items in the original list, the programme includes low taxes, a minimal state that denies having any responsibility for income distribution, either a currency board or freely floating exchange rates, and rapid liberalization of the capital account. It has even been suggested that the Washington Consensus implies a belief that all markets are perfectly competitive so that neoclassical economic analysis is literally true always and everywhere and government action is always a mistake (Stiglitz 2006, p. 24). Despite the name, those who used the term in this way regarded it as unnecessary to establish that these attitudes characterized large parts of Washington. Personally I do not recall having met anyone in Washington who subscribes to these rather bizarre beliefs.

The most controversial elements of the Washington Consensus proved to concern microeconomic liberalization. Points 4 and 6–9 of my original list pointed in that direction, as did especially the last alternative interpretation of the phrase alluded to above. In the wake of the collapse of communism I overestimated the extent to which the desire for an active government role in managing the economy would be replaced by the acceptance of market mechanisms. The experience of East Asia was frequently invoked to question whether a commitment to laissez-faire had really been the key to the region's fast growth. On the contrary, many critics of the Washington Consensus argued that the success of the East Asian countries was due to their governments having implemented industrial policies. This runs into the objection that at least one of the East Asian success stories (Hong Kong) had developed under the closest to a laissez-faire system that the world has ever seen. Why should one attribute the success of Korea and Taiwan to what distinguishes them from Hong Kong and to some extent Singapore rather than to what they have in common (for example, macro stability, high savings, export orientation, good education and an early demographic transition)? The old questions of market versus state, or even the merits of industrial policy, have not been settled by these arguments.

See Also

► Globalization

Bibliography

Burki, J., and G.E. Perry. 1998. *Beyond the Washington consensus: Institutions matter*. Washington, DC: World Bank.

Kuczynski, P.-P., and J. Williamson. 2003. *After the Washington consensus: Restarting growth and reform in Latin America*. Washington, DC: Institute for International Economics.

Naím, M. 1994. Latin America: The second stage of reform. *Journal of Democracy* 5: 32–48.

Stiglitz, J.E. 2006. *Making globalization work*. New York: Norton.

Williamson, J. 1990. What Washington means by policy reform. In *Latin American adjustment: How much has happened?* ed. J. Williamson. Washington, DC: Institute for International Economics.

Water Resources

John V. Krutilla

Interest in the economics of water resources had its inception in the United States upon the passage of the Flood Control Act of 1936. That legislation

specified that '... benefits to whomsoever they may accrue, exceed the costs' to justify project development. That statement both implied an efficiency criterion and anticipated the Hicks–Kaldor compensation principle. Thereafter, the newly created Subcommittee on Benefits and Costs of the Interagency River Basin Committee began to elaborate the economic implications of this section of the legislation. Attention was initially focused on getting an appropriate investment criterion, which addressed not only the issues of optimally designing, sizing and timing projects, but also questions of estimating the value of non-priced services such as flood damage reduction and the value of a user-free waterway for barge traffic. In addressing these issues the Subcommittee on Benefits and Costs had to deal practically with several issues that had attracted, or were to attract, the attention of the economics profession.

One of these was to the measurement of the contribution of a project to consumer well-being, when services are not marketed or outputs are large relative to a market. On the larger rivers the site for an impoundment, if developed to a scale that would minimize unit costs, often produced a facility having an output that was large in relation to the existing market (for the the pre-World War II United States economy) not unlike current circumstances in developing countries. Making up a large share of total system capacity, the output of a hydroelectric facility, for example, would have to be sold at a substantial reduction in pre-project price in order to clear the market. Whether from pragmatic considerations, or simple intuition, the Subcommittee recommended that the block of energy be valued at neither the price that ruled before the event, nor at the new market-clearing price, but rather by a price midway between the pre- and postproject prices. Now if the demand for power could be taken to be approximated by a linear function, we observe that this is virtually equal to the area under the demand curve, an accurate measure of the willingness to pay for the service. This issue was treated more precisely in Eckstein's later work (1958).

Another issue addressed by the Subcommittee was the problem of estimating the value of project outputs which are public goods. It is a characteristic of flood management using reservoir storages, that if the system is managed to provide flow regulating services for one occupant in the flood plain, it simultaneously provides damage reduction services for all occupants in the flood plain. And the value of the damage reduction enjoyed by any occupant does not diminish the value of the service to other flood plain occupants. In short, flood stage reducing services of storage reservoirs have the characteristics of a public good. This was discovered as a practical matter before the economic treatment of the problem by Samuelson (1954). In spite of the potential difficulties stemming from non-revelation of preferences for public goods, an estimate of flood control benefits was obtained indirectly by estimates of the demand for substitute market goods in the manner detailed by Karl-Göran Mäler (1972) That is, given the hydrologists' estimates of the recurrence interval of various flood stages with and without the flood control projects, and the difference in damages to facilities in the flood plain that would occur under two regimes, an estimate of the value of the flood management services could be obtained. This procedure was employed to estimate the value of flood control facilities of the Tennessee Valley Authority throughout the 1930s and by the US Corps of Engineers in the remainder of the country.

In addition to awareness of project valuation problems associated with indivisible inputs and outputs, direct or physical interdependence among facilities, a pervasive phenomenon in the natural resource sector, was also recognized early. The value of downstream hydroelectric facilities also was known to be linked to the presence of upstream storage and thus direct interdependence received early attention in planning and valuing water resource systems. This is additional to the wide variety of measurable externalities which were treated routinely in river basin planning.

The conceptual framework that supported the work carried on during the 1930s and 1940s was largely the contribution of the planning engineers assisted by the work of the Subcommittee on Benefits and Costs. Where irrigation agriculture was involved, as it was in all of the arid western

states, agricultural economists with the United States Department of Agriculture (USDA) Bureau of Agricultural Economics (BAE) (now the Agricultural Research Service) were notably present. Perhaps the most prominent for his contribution to this area was Mark Regan. The dean of the academic agricultural economists whose contribution was substantial was Professor S.V. Ciriacy-Wantrup of the University of California, Berkeley. It should be mentioned that during the 20 years following the Flood Control Act of 1936, there was little interest in the economics of water resources by academic economists, except for a small number in the agricultural colleges in the western United States where irrigation agriculture was extensively practised.

During the mid- to late 1950s, however, the environment changed dramatically. The water resource ventures in the 1930s and 1940s were an expression of President Roosevelt's New Deal philosophy. In 1952 there was the first change in political parties to head up the national administration in 20 years. The new administration had a pronounced tilt toward using the market and private ventures to substitute for practically all public ventures. Along with this change in policy went a reduction in budgets for public sector expenditures. This change of emphasis in Washington created an environment in which both public and private development had to be justified. The new emphases of this changed environment stimulated a number of economic inquiries by academic economists.

At the end of the 1950s, three studies with somewhat different emphases emerged almost simultaneously. One, by Otto Eckstein (1958), addressed the subject matter of the Subcommittee on Benefits and Costs (1950), providing a more rigorous theoretical structure for the evaluation of water resource projects, including investment criteria to cover the constrained budget case. A second study by Krutilla and Eckstein (1958) addressed the question of the relative efficiency of public and private development of water resource systems, identifying cases in which, under the then current institutional arrangement (Federal Power Act and related legislation) each alternative had the advantage. While this study was undertaken in the spirit of the Hicks–Kaldor compensation criterion it nonetheless undertook the first significant study of the distributional aspects of a public works expenditure. The third volume by McKean (1958) attempted to bring advances in weapon systems evaluation methodology developed at the Rand Corporation to investment analysis in the water resources field. All of these studies, without being explicit, nonetheless relied heavily on old fashioned neoclassical welfare economics.

Perhaps the most comprehensive single study of water resource economics was provided by Hirshleifer et al. (1960). This study addressed the issue of more efficient allocation of existing supplies as one option to more extensive development of additional sources. The volume, incidentally, represented a thorough-going critique of water resource policies, laws and institutions.

The terminal study of this generation of water studies is the voluminous 'report' of the Harvard Water Resource Seminar, a multi-year seminar that was conducted by members of the Harvard political science, economics and engineering faculties and hosted among its students, practicing public sector professionals regularly engaged in various roles in public water resource agencies.

The combination of analytic vigour and field experience present in the Harvard study served to reveal the interstices between theory and application and make more evident the significance of distributional considerations that constrain the politically feasible optimization prescriptions. The concern with distributional considerations was not confined to the deliberations of the Harvard Water Resources Seminar (Maass et al. 1962). Krutilla and Eckstein (1958) had already addressed the distributional consequences of public expenditures. This was done, however, as a descriptive, rather than as an explicit policy analytic, exercise. The Harvard seminar, the work of Haveman (1965), Freeman (1967), Haveman and Krutilla (1968) and Tolley (1959) investigated the distributional consequences as policy issues bringing up more prominently the significance of distributional considerations than had been done previously.

As the water resource economics field matured the nature of the studies tended to specialize to the

various individual functions that water resources development provides. A partial accounting of these would include: for irrigated agriculture, Ruttan (1965), Frederick (1975), Crosson, Cummings, and Frederick (1978); for interbasin transfers, Howe and Easter (1971), Cummings (1974), Hartman and Seastone (1975); for hydropower and flood control development on an international stream, Krutilla (1967) and Lind (1967); and for inland waterway transportation, Howe and associates (1969).

Another study carried out in the same spirit and addressing qualitative aspects of streamflows was the seminal work by Allen Kneese (1964) which brought a whole new dimension to water resources research. It introduced to economic analysis the field of environmental economics, and indeed most of the work in water resources for the next two decades addressed environmental issues of various kinds.

The general field of water resources appeared to many to offer important opportunities to apply welfare economics to practical problems. In this regard, while many of the studies contented themselves with meeting the Hicks–Kaldor compensation criterion, others attempted to implement Little's dictum that the income redistribution of such public undertakings be 'good' (1950). It may be wondered how the present essay can be written without reference to Little's own work applying many of the same analytical techniques in a somewhat similar environment – industrial investment projects in the developing countries. But the work of Little and Mirrlees (1968, 1969) was not specifically related to investments in water resource projects, nor was the work of Mishan (1972) and Squire and Van der Tak (1975). Nevertheless, they all addressed a common genre of problems in an environment where new capacity involved lumpiness, where the outputs involved public goods in part and where externalities similarly bedevilled market indices of value.

See Also

- ► Natural Resources
- ► User Fees

Bibliography

Ciriacy-Wantrup, S.V. 1955. Benefit-cost analysis and public resource development. *Journal of Farm Economics* 37(4), November, 676–689.

Ciriacy-Wantrup, S.V. 1964. Concepts used as economic criteria for a system of water rights. In *Economics and public policy in water resources development*, ed. S.C. Smith and E.N. Castle. Ames: Iowa State University Press.

Crosson, P.R., R.C. Cummings, and K.D. Frederick (eds.). 1978. *Selected water management issues in Latin American agriculture*. Baltimore/London: Johns Hopkins University Press for Resources for the Future.

Cummings, R.G. 1974. *Interbasin water transfers: A case study in Mexico*. Washington, DC: Resources for the Future.

Eckstein, O. 1958. *Water resources development: The economics of project evaluation*. Cambridge, MA: Harvard University Press.

Frederick, K.D. 1975. *Water management and agricultural development: A case study of the Cuyo Region of Argentina*. Baltimore: Johns Hopkins Press for Resources for the Future.

Freeman, A.M. 1967. Income distribution and public investment. *American Economic Review* 57(3, June): 495–508.

Hartman, L.M., and D. Seastone. 1975. *Water transfers, economic efficiency and alternative institutions*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Haveman, R.H. 1965. *Water resource investment and the public interest*. Nashville: Vanderbilt University Press.

Haveman, R., and J.V. Krutilla. 1968. *Unemployment, idle capacity, and the valuation of public expenditures: National and regional analyses*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Hirshleifer, J., J.C. Dehaven, and J.W. Milliman. 1960. *Water supply economics, technology and policy*. Chicago: University of Chicago Press.

Howe, C.W., and K.W. Easter. 1971. *Interbasin transfer of water: Economic issues and impacts*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Howe, C.W., J.L. Carroll, A.P. Hurter Jr., W.J. Leininger, S.G. Ramsey, N.L. Schwartz, E.A. Silberberg, and R.M. Steinberg. 1969. *Inland waterway transportation: Studies in public and private management and investment decisions*. Washington, DC: Resources for the Future.

Kneese, V. 1964. *The economics of regional water quality management*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Krutilla, J.V. 1961. Welfare aspects of benefit-cost analysis. *Journal of Political Economy* 69(3, June): 226–235.

Krutilla, J.V. 1967. *The Columbia river treaty: The economics of an international river basin development*. Baltimore: Johns Hopkins University Press for Resources for the Future.

Krutilla, J.V., and O. Eckstein. 1958. *Multiple purpose river development: Studies in applied economic analysis*. Baltimore: Johns Hopkins University Press for Resources for the Future.
Lind, R.C. 1967. Flood control alternatives and economics of flood protection. *Water Resources Research* 3(2): 345.
Little, I.M.D. 1950. *A critique of welfare economics*. Cambridge: Cambridge University Press.
Little, I.M.D., and J.A. Mirrlees. 1968. *Manual of industrial project analysis in developing countries*, vols I and II. Paris: Development Centre of the Organization of Economic Co-operation and Development.
Maass, A., M.H. Hufschmidt, R. Dorfman, H.A. Thomas Jr., S. Marglin, and G.M. Fair. 1962. *The design of water-resources systems*. Cambridge, MA: Harvard University Press.
Mäler, K.-G. 1972. *Environmental economics*. Baltimore: Johns Hopkins University Press for Resources for the Future.
McKean, R.N. 1958. *Efficiency in government through systems analysis, with emphasis on water resource development*. New York: Wiley.
Mishan, E.J. 1972. *Economics for social decisions; elements of cost-benefit analysis*. London: George Allen & Unwin.
Ruttan, V.W. 1965. *The economic demand for irrigated acreage: New methodology and some preliminary results, 1954–1980*. Baltimore: Johns Hopkins University Press for Resources for the Future.
Samuelson, P.A. 1954, November. The pure theory of public expenditures. *Review of Economics and Statistics* 36: 387–389.
Squire, L., and H.G. Van der Tak. 1975. *Economic analysis of projects*. Baltimore: Johns Hopkins University Press for the World Bank.
Steiner, P.O. 1959. Choosing among alternative public investments in the water resources field. *American Economic Review* 49(5, December): 893–916.
Subcommittee on Benefits and Costs. 1950. *Report to the* Federal Interagency River Basin Committee, *Proposed practices for economic analysis of river basin projects*.
Tolley, G. 1959. Reclamations influence on the rest of agriculture. *Land Economics* 35(2, May): 176–180.

Wavelets

James B. Ramsey

Abstract

Wavelets provide a flexible basis for representing a signal that can be regarded as a generalization of Fourier analysis to non-stationary processes, or as a filter bank that can represent complex functions that might include abrupt changes in functional form, or signals with time varying frequency and amplitude. Of greatest import for economic analysis is the orthogonal deconstruction of a signal into time scale components that allow economic relationships to be analysed time scale by time scale and then re-synthesized.

Keywords

Capital asset pricing model; Denoising; Discrete wavelet transform; Filter banks; Forecasting; Fourier analysis; Fourier transforms; Heisenberg uncertainty principle; Multiresolution analysis; Projection pursuit; Regime shifts; Serial correlation; Shrinkage; Threshold models; Waveform dictionaries; Wavelet packets; Wavelets

JEL Classifications

C32; C33; C43; C5

Wavelets provide a powerful tool of analysis for economics and finance, as well as for scientists in a wide variety of fields, such as signal processing, medical imaging, data compression and geology. One interpretation of wavelets is that they are a collection of functions that provide a basis for the representation of test functions that may be complicated with localized shocks, have abrupt changes in functional form, or are signals with time varying frequency and amplitude.

Another interpretation is that wavelets are a generalization of Fourier analysis in which stationarity of the time series is no longer critical and localization of a signal can be achieved. In this light, to borrow a great insight by Strang (see Strang and Nguyen 1996), Fourier analysis is best at representing functions that are composed of linear combinations of stationary inputs, but wavelets are like musical notation in that each note is characterized by its frequency, its position in time, and its duration.

A further interpretation of wavelets is that of a filter bank, so that different classes of wavelet functions are generated by prescribing different banks of filters. Filter banks can achieve results that are not possible with a single filter (Strang and Nguyen 1996). Yet another interpretation is that of a decomposition of a signal in terms of different time scales, an interpretation that is at the heart of much economic analysis as represented by the long-standing notions of the 'short, medium and long runs' and is fundamental to the concept of the 'business cycle'.

At this time there is a vast literature on wavelets in mathematics, statistics, and various branches of engineering, but relatively little in economics, although that situation is changing fast. An introduction to the economic literature that is highly recommended is Gencay et al. (2002). This is the most comprehensive and detailed coverage with numerous descriptions of economic applications and discussions of the statistical properties of the wavelet estimators. Bruce and Gao (1996) discuss the properties of wavelets and give instructions for calculating wavelets using S-Plus; Chui (1992), Percival and Walden (2000), and Strang and Nguyen (1996) develop the mathematics at a moderate level of difficulty and discuss the statistical properties of wavelets. Silverman and Vassilicos (1999) provide interesting examples of the applications of wavelets and further discussions of the statistical properties (see also Ramsey 1999a). Two lower-level introductions for economists are Crowley (2005) and Schleicher (2002).

Two Informative Examples

Let **X** represent an N dimensional vector of observations on a time series or a function $f(\cdot)$ evaluated at the discrete points $t = 1,2,\ldots N$. One may consider an orthonormal transformation from **X** to an N dimensional vector **W**, with elements, W_n, $n = 0,1,\ldots N-1$, where **W** is generated from an $N \times N$ dimensional orthonormal matrix W; $\mathbf{W} = W\mathbf{X}$ and $I_N = W^T W$. Let $N = 2^J$ for some integer J. This assumption, while inessential, is analytically convenient and is useful for defining an efficient algorithm for evaluating the wavelet coefficients. The N^2 elements of the transformation matrix W are the filter elements to be defined below and the N elements of the vector **W** are the wavelet coefficients that are given by the inner product $W\mathbf{X}$.

The wavelet defined here is known as the discrete wavelet transform (DWT). By choosing an orthonormal transform it is immediate that the modulus of **W** is the same as that of **X**; that is, $\|\mathbf{X}^2\| = \|\mathbf{W}^2\|$, where $\|\mathbf{X}^2\|$ is given by $\mathbf{X}^T\mathbf{X} = \sum x_i^2$, and where w_n^2 is the energy contributed by the n^{th} wavelet coefficient.

The power of wavelet analysis lies in the choices made for the components of W. Two examples illustrate the choices that can be made and indicate the scope that a wavelet analysis offers. (I am indebted to Percival and Walden 2000, for the examples to follow.) The first example is the oldest and simplest function used to generate a wavelet transformation. For the sake of simplicity of exposition, choose $N = 2^J = 2^4 = 16$. The notation, $W_{i,*}^T$ indicates the i^{th} row of the matrix W, $i = 0,1,\ldots N-1$. The function used to generate the elements of W for the DWT is the Haar function.

$$
\begin{aligned}
W_{0,*}^T &= \left\{-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0_1, \ldots 0_{14}\right\},\\
W_{8,*}^T &= \left\{-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0_1, \ldots 0_{12}\right\}\\
W_{12,*}^T &\\
&= \left\{-\frac{1}{\sqrt{8_1}}, \ldots -\frac{1}{\sqrt{8_4}}, \frac{1}{\sqrt{8_1}}, \ldots \frac{1}{\sqrt{8_4}}, 0_1, \ldots 0_8\right\}\\
W_{14,*}^T &= \left\{-\frac{1}{4_1}, \ldots -\frac{1}{4_8}, \frac{1}{4_1}, \ldots \frac{1}{4_8}, \right\},\\
W_{15,*}^T &= \left\{\frac{1}{4_1}, \ldots \frac{1}{4_{16}}, \right\}
\end{aligned}
\tag{1}
$$

The unshifted filters are $W_{0,*}^T, W_{8,*}^T, W_{12,*}^T, W_{14,*}^T, W_{15,*}^T$. The remaining rows of W, are circularly shifted versions of the filter terms defined above. We have:

$$
\begin{aligned}
W^T_{1,*} &= T^2\, W^T_{0,*} = \left\{0,0,\ -\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}, 0_1, \ldots 0_{12}\right\} \\
W^T_{2,*} &= T^4\, W^T_{0,*} = \left\{0,0,0,0,\ -\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}, 0_1, \ldots 0_{10}\right\} \\
W^T_{9,*} &= T^4\, W^T_{8,*} = \left\{0,0,0,0,\ -\frac{1}{2},\ -\frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}, 0_1, \ldots 0_8\right\} \\
W^T_{13,*} &= T^8\, W^T_{12,*} = \left\{0_1, \ldots 0_8,\ -\frac{1}{\sqrt{8_1}}, \ldots -\frac{1}{\sqrt{8_4}}, \frac{1}{\sqrt{8_1}}, \ldots \frac{1}{\sqrt{8_4}}\right\}
\end{aligned}
\tag{2}
$$

The filter $W^T_{15,*}$ yields the 'scaling coefficient', and the remaining filters, $W^T_{i,*}$ yield the wavelet coefficients. From the above equations one sees that the Haar filters involve first differences between scale averages at non-overlapping intervals. At the lowest scale, 2^{j-1} for $j = 1$, the first differences are between adjacent observations, at the next scale level, 2^{j-1} for $j = 2$, the differences are between adjacent pairs of observations, at the scale 2^{j-1} for $j = 3$ the differences are between groups of four terms, and so on. At the highest scale, $J = 4$, we have two filters of the data, a first difference between the first and last eight of the observations, and an average over the full set of observations. The latter is a 'father' wavelet transform and the remaining rows of W are the 'mother' wavelet transforms.

In order to gain further insight, consider a Daubchies wavelet designated as D (4) (see Daubchies 1992), which is a member of a sequence of discrete wavelet filters in which the Haar is D(2). Define $Y_t \equiv aX_t + bX_{t-1}$ and form the backward second discrete difference, $Y^{(2)}_t$ by:

$$
Y^{(2)}_t \equiv Y^{(1)}_t - Y^{(1)}_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}. \tag{3}
$$

For a particular choice of 'a' and 'b', the nth D (4) wavelet can be written as:

$$
\begin{aligned}
W_n &= Y_{2n+1} - 2Y_{2n} + Y_{2n-1} \\
&= aX_{2n+1} + (b-2a)X_{2n} + (a-2b)X_{2n-1} \\
&\quad + bX_{2n-2} \\
&= h_0X_{2n+1} + h_1X_{2n} + h_2X_{2n-1} + h_3X_{2n-2}.
\end{aligned}
\tag{4}
$$

Imposing orthonormality {and some other conditions to ensure uniqueness}, enables one to derive the values for h_i, i = 0, 1, 2, 3, namely:

$$
h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}}, h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}}, h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}},
$$

$$
h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}.
$$

Repeating the exercise above with $N = 2^J = 16$ observations, we have:

$$
\begin{aligned}
W^T_{0,*} &= \{h_1, h_0, 0_1, \ldots 0_{12}, h_3, h_2\}, \\
W^T_{1,*} &= \{h_3, h_2, h_1, h_0, 0_1, \ldots 0_{12},\}
\end{aligned}
$$

where $W^T_{1,*} = \mathrm{T}^2 W^T_{0,*}$, orthonormality requires that $||W_{0,*}||^2 = h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1$ and $\left\langle W^T_{0,*} W_{1,*} \right\rangle = h_0h_2 + h_1h_3 = 0$. $W^T_{15,*}$ applied to the time series yields the mean.

A wavelet filter of length L must satisfy at a minimum for all nonzero integers n the following conditions.

$$
\sum_{l=0}^{L-1} h_l = 0, \quad \sum_{l=0}^{L-1} h_l^2 = 1 \tag{5}
$$

$$
\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0.
$$

For the Haar wavelet, D(2), and the Daubchies D(4) wavelet L is respectively 2 and 4.

Wavelets and Multiresolution Analysis (MRA)

One can approach the definition of wavelets from a related perspective that indicates the similarities

to and differences from Fourier transforms. In both cases, one is considering a projection of a signal onto a set of basis functions for the space containing the test function. In the case of the Fourier transform the basis functions are rescalings of the fundamental frequency, for example, $\{e^{-in\omega_0}\}$ where ω_0 is the fundamental frequency and n provides the scaling. In this expression there is no resolution in the time domain, and the signal is assumed to be stationary. In contrast, the wavelet-generating functions are defined over very general spaces and each function is compact. One has the recursive relationship:

$$g_{s,k}(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-k}{s}\right) \tag{6}$$

where s indicates the scale of the function, k is the location index; the term $\sqrt{s}$ ensures that the norm of $g(\cdot)$ is 1. The projection of the signal onto the scalable function $g(\cdot)$ depends on two parameters; s, which defines the time scale and implicitly designates a relevant range of frequencies, and k, which indicates the centre of location of the projection. The compactness of $g(\cdot)$ together with the time index k implies that the analysis of a time series is essentially local at each scale, whereas the Fourier analysis is essentially global. There is considerable latitude in the choice of the function $g(\cdot)$, or in the choice of the filters that generate the functions $g(\cdot)$. Desirable criteria include symmetry, smoothness and orthogonality. Whether one begins by specifying the properties of a basis function $g(\cdot)$ or one begins by specifying the properties of a filter $\{h_i\}$, the process generates two related classes of wavelet transforms, the 'father' wavelet that yields the 'scaling' coefficients and the 'mother' wavelets that yield the detail coefficients.

One can link the filter coefficients to the definition of the father and mother wavelets and link the father and mother transforms themselves by noting that, for a given sequence of low pass filter coefficients, $l(k)$, and the corresponding high pass filter coefficients, $h(k)$, one solves for $\Phi(t)$, father, and $\varphi(t)$, mother, from:

$$\Phi(t) = \sqrt{2}\sum_{k=0}^{N} l(k)\Phi(2t-k) \tag{7}$$

$$\Psi(t) = \sqrt{2}\sum_{k=0}^{N} h(k)\Phi(2t-k).$$

For the Haar example above, the filter coefficients are: $l(k) = \left\{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right\}$ and $h(k) = \left\{-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right\}$.

From these equations, one derives the scaling, or 'smooth' coefficients and the wavelet or 'detail' coefficients of the function $f(\cdot)$ by the integrals,

$$s_{J,k} = \int f(t)\Phi_{J,k}dt \tag{8}$$

$$d_{j,k} = \int f(t)\Psi_{j,k}dt, j = 1, 2, \ldots J \tag{9}$$

where $\Phi_{J.k}$ and $\varphi_{J,k}$ are the scaled and translated versions of Φ and φ defined in Eq. (7). The function $f(t)$ can be synthesized by the equations:

$$\begin{aligned} f(t) &= \sum_k s_{J,k}\Phi_{J,k}(t) + \sum_k d_{J,k}\Psi_{J,k}(t) \\ &\quad + \ldots \sum_k d_{j,k}\Psi_{j,k}(t) \ldots + \sum_k d_{1,k}\Psi_{1,k}(t) \\ f(t) &= S_J + D_J + D_{J-1} + \ldots D_j + \ldots D_1 \\ S_J &= \sum_k s_{J,k}\Phi_{J,k},\ D_j = \sum_k d_{j,k}\Psi_{j,k}, j = 1,2,\ldots J \\ f(t) &\approx S_{j-1} = S_j + D_j. \end{aligned} \tag{10}$$

An easy way to visualize the above scale and locational decomposition of the signal is as a series of maps of ever greater detail as elements of D_j are added; S_J provides a smooth outline, D_J adds the highest scale detail; and the D_j add ever more detail as j decreases. One can approximate the function $f(t)$ by truncating the expansion at some j, $1 < j < J$. This is known as a multiresolution analysis (MRA), which can yield enormous data compression by representing the function $f(t)$ with relatively few coefficients.

There are many choices for the basis function, or alternatively choices for the filter banks that provide a great deal of richness for the wavelet approach. Two generalizations of note are wavelet packets and an exploratory technique known as waveform dictionaries. The former analyses signals by basis functions that differ by location and scale as for wavelets, but also by an oscillation index; wavelet packets are most useful in representing time series that have short term, localized oscillations (Bruce and Gao 1996). Waveform dictionaries (Mallat and Zhang 1993) provide a modification to wavelet analysis. The basic function providing the basis is a function g (.) defined by:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{i\omega t}, \gamma = (s, u, \omega). \quad (11)$$

The function $g_\gamma(t)$ has norm one, has scale s, and the time scale energy is centred at u and proportional to s. The Fourier transform of $g_\gamma(t)$ has its frequency energy centred at ω and is proportional to $1/s$. The dictionary of functions $g_\gamma(t)$ illustrate a very important principle of these transforms. Improved resolution in the time domain reduces resolution in the frequency domain and vice versa; this is a version of the Heisenberg uncertainty principle.

Applications in Economics and Finance

All analytical procedures can be assessed on the basis of their contribution in four categories: provide estimators in novel situations; improve efficiency or reduce bias; enhance robustness to modelling errors; or provide new insights into the data-generating processes. Wavelets have provided benefits in all these categories.

One advantage of the waveform dictionary approach indicated in Eq. (11) is that the researcher need not prejudge the presence of frequency components as well as the occurrence of short-term shocks. The process is exploratory and projection pursuit methods can be utilized to isolate local and global characteristics. Waveform dictionaries have been used as an exploratory tool in the analysis of financial and foreign exchange data (Ramsey and Zhang 1996, 1997). In the analysis of daily stock-price data and tic-by-tic exchange rate data, there was no strong structural evidence for any frequency, but there was weak evidence for frequencies that appeared and disappeared or that waxed and waned in strength. Most of the power was summarized in terms of time-localized bursts of activity. The results in both papers indicated that, while for any given time period surprisingly few wavelet coefficients were needed to fit the data, the relevant coefficients varied randomly from period to period. Each burst was characterized by a rapid increase in amplitude and fast oscillation in frequency; in short, market adjustment processes seem to be characterized by a rapid increase in oscillation amplitude and frequency followed by a decay in frequency and amplitude; adjustment is neither smooth nor fast.

For a deep analysis of the scaling properties of volatilities and the relationship between risk and time scales, see Gencay et al. (2001, 2003). Another example in these references is the estimation of time varying Betas in the capital asset pricing model (CAPM). The analysis indicated that in the cases examined beta coefficients varied substantially over time, thereby modifying the structure of optimal investment strategies.

Wavelets have been instrumental in improving the robustness and efficiency of estimation in numerous examples (see Jensen 1999, 2000) for efficiency gains and enhanced robustness of estimates for the fractional differencing parameter in long memory processes (see also Gencay et al. 2002). This reference is also useful for examples of estimation of covariance matrices and providing confidence intervals.

Wavelets have been successfully employed in situations not amenable to standard approaches – for example, forming estimators in testing for serial correlation of unknown form in panel models. As Hong and Kao (2004) state in their abstract: 'This paper proposes a new class of generally applicable wavelet-based tests for serial correlation of unknown form in the estimated residuals of a panel regression model, where error components can be one-way or two-way, individual and time effects can be fixed or random, and regressors may

contain lagged dependent variables or deterministic/stochastic trending variables.'

Ramsey and Lampart (1998a, b) discovered that the relationship between economic variables – for example, between money and income, or between consumption and income – can be decomposed into relationships at separate scales with interesting implications for economic theory. The insights gained include reconciling a number of anomalies in the consumer and monetary theory literature as well as introducing the notion that observed time series are an aggregation of behaviour at different time scales; for example, the time path of consumption totals represents the actions by consumers operating on a variety of time scales. In the same papers the authors also discovered that at certain time scales the relationship between economic variables may be subject to variations in the delay. These results have been confirmed by other researchers (for example, Gencay et al. 2002). For an alternative approach to testing for causality in the frequency domain using wavelets, see Dalkir (2004).

Yet another insight provided by wavelet analysis is the distinction between 'smoothing' and denoising. The former, traditional in econometrics, is based on the assumption that the signal is smooth relative to the noise, whereas the latter allows for the signal to be as irregular as the noise, but with greater amplitude. For smooth signals subject to noise, the obvious approach in order to minimize the effect of the noise is to average in some manner. However, if the signal is not smooth, averaging is not a suitable approach in that the averaging process distorts the signal itself. One can claim that denoising is often more relevant to economic and financial analysis than is smoothing (see Ramsey 2004). These remarks are particularly relevant in the context of estimating relationships involving regime shifts, threshold models, and other non-differential changes in variable values. In an important series of papers Donoho, Johnstone and coauthors explored the use of wavelets and the concept of shrinkage whereby the size of the wavelet coefficient estimates is reduced to allow for the presence of noise (see Donoho et al. 1995). Further, shrinkage can be applied differentially across scales thereby refining the technique (see Ramsey 2004, for more recent references and Gencay et al. 2002, for a thorough development of wavelet denoising).

Forecasting is an important topic; see Fryzlewicz et al. (2002) and Li and Hinich (2002), who demonstrate how the wavelet approach disentangles the variation in forecastability over time scales; that is, the ability to forecast varies across time scales. At the simplest level a given time series can be decomposed into trend, business cycle and seasonal components by wavelets and individually structured forecasting methods applied to each component separately before synthesizing the entire signal in order to produce forecasts for the whole series (see Ramsey 2004, for a brief review of the literature on forecasting using wavelets).

See Also

- Forecasting
- Long Memory Models
- Seasonal Adjustment
- Spectral Analysis
- Structural Change
- Time Series Analysis

Bibliography

Brillinger, D.R. 1994. Some uses of cumulants in wavelet analysis. *Journal of Nonparametric Statistics* 6: 93–114.

Bruce, A., and H. Gao. 1996. *Applied wavelet analysis with S-plus*. New York: Springer.

Chui, C.K. 1992. *An introduction to wavelets*. New York: Academic Press.

Crowley, P. 2005. An intuitive guide to wavelets for economists. Discussion Paper No. 1/2005, Research Department, Bank of Finland.

Dalkir, M. 2004. A new approach to causality in the frequency domain. *Economics Bulletin* 3(44): 1–14.

Daubchies, I. 1992. *Ten lectures on wavelets*. Philadelphia: SIAM.

Donoho, D.L., I.M. Johnstone, G. Keryachariam, and D. Picard. 1995. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society* 57: 301–369.

Fryzlewicz, P., van Bellegem, S. and von Sachs, R. 2002. Forecasting non-stationary time series by wavelet process modeling. Technical Report No. 0208, Institut de Statistique, Université Catholique de Louvain.

Gencay, R., F. Selcut, and B. Whitcher. 2001. Scaling properties of foreign exchange volatility. *Physica A* 289: 249–266.
Gencay, R., F. Selcut, and B. Whitcher. *2002*. *An introduction to wavelets and other filtering methods in finance and economics*. *New York*: *Academic Press*.
Gencay, R., F. Selcut, and B. Whitcher. 2003. Systematic risk and time scales. *Quantitative Finance* 3: 108–116.
Hong, Y., and C. Kao. 2004. Wavelet testing for serial correlation of unknown form in panel models. *Econometrica* 72: 1519–1563.
Jensen, M.J. 1999. Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting* 18: 17–32.
Jensen, M.J. 2000. An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control* 24: 361–387.
Li, T.H., and M. Hinich. 2002. A filter bank approach for modeling and forecasting seasonal patterns. *Technometrics* 44(1): 1–14.
Mallat, S., and Z. Zhang. 1993. Matching pursuit with time frequency dictionaries. *IEEE Transactions on Signal Processing* 41: 3397–3415.
Percival, D.B., and A.T. Walden. 2000. *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press.
Ramsey, J.B. 1999a. The contribution of wavelets to the analysis of economic and financial data. In *Philosophical transactions of the royal society, series A: Mathematical and physical sciences.* Repr. in *Wavelets,* ed. B.W. Silverman and J.C. Vassilicos. Oxford: Oxford University Press.
Ramsey, J.B. 1999b. Regression over time scale decompositions: A sampling analysis of distributional properties. *Economic Systems Research* 11(2): 163–183.
Ramsey, J.B. 2004. Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics and Econometrics* 6(3): Article 1.
Ramsey, J.B., and C. Lampart. 1998a. The decomposition of economic relationships by time scale using wavelets: Money and income. *Macroeconomic Dynamics* 2: 49–71.
Ramsey, J.B., and C. Lampart. 1998b. The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies in Nonlinear Dynamics and Econometrics* 3(4): 23–42.
Ramsey, J.B., and Z. Zhang. 1996. The application of waveform dictionaries to stock market index data. In *Predictability of complex dynamical systems*, ed. Y.-A. Kravtsov and J. Kadtke. New York: Springer.
Ramsey, J.B., and Z. Zhang. 1997. The analysis of foreign exchange rates using waveform dictionaries. *Journal of Empirical Finance* 4: 341–372.
Schleicher, C. 2002. An introduction to wavelets for economists. Working Paper No. 2002-3, Bank of Canada.
Silverman, B.W., and J.C. Vassilicos. 1999. *Wavelets: The key to intermittent information?* Oxford: Oxford University Press.
Strang, G., and T. Nguyen. 1996. *Wavelets and filter banks*. Wellesley: Wellesley-Cambridge Press.

Wealth

Robert L. Heilbroner

Keywords

Abstract labour; Circulating capital; Cost of production; Exchange value; Labour; Marginal revolution; Power; Productive and unproductive labour; Property; Riches; Scarcity; Social stratification; Stocks and flows; Utility; Value; Wage differentials; Wealth

JEL Classifications

A1

Wealth is a fundamental concept in economics – indeed, perhaps the conceptual starting point for the discipline. Despite its centrality, however, the concept of wealth has never been a matter of general consensus. Although wealth has not become a focus of heated controversy comparable to that of value (despite the fact that the two terms are inextricably conjoined, as we shall see), conceptions of wealth have clashed profoundly and even irreconcilably. The result has been a continuing discussion of deep importance for economics – not only for its intrinsic interest, but because it calls into question the very scope and content of the discipline itself.

At the root of the long history of disagreement about wealth lie two conflicting conceptions of what the word implies. One of these, far more ancient than the formal study of economics and still very much in general use, is the idea of wealth as tangible possessions. For over a century, however, this conception has been challenged by another, which has identified the nature of wealth in the pleasures or 'utilities' generated by tangible goods, rather than the goods themselves. In the differing implications arising from these 'objective' and 'subjective' conceptions of wealth lie consequences of great significance for a discipline that has traditionally considered itself to be concerned with the study of wealth.

The objective conception of wealth is as old as written history, but the economist has not been interested in records of slaves and land and gold, other than to remark (usually as an economic anthropologist) on the extraordinary variety of objects that have been utilized as embodiments of wealth. The analytic problem to which economists have been drawn has been the attempt to establish a common denominator in which to sum up the value represented by a heterogenous collection of objects. 'The entire study of wealth is, indeed, meaningless unless there be a unit for measuring it;' wrote J.B. Clark, 'for the questions to be answered are quantitative. How great is the wealth of a nation?' (Clark 1899, p. 375).

In ordinary discourse, this common denominator has always been money, and we will later consider the cogency of this common sense rule. For the economist, however, the challenge has been to discover some metric less arbitrary and unstable than a monetary sum. Thus the idea of objective wealth becomes inextricably entwined with the need to discover a standard – an embodiment of 'value' – by which its extent can be calculated. In the late mercantilist period this measure of extent was conceived by Petty and Cantillon to be the 'amounts' of land and labour that entered into the production of things – a considerable advance over earlier ideas that gold and silver possessed intrinsic value. This dual standard was subsequently reduced by Adam Smith to labour alone. 'It was not by gold or silver, but by labour that all the wealth of the world was originally purchased', he wrote in *The Wealth of Nations,* 'and its value, to those who possess it and who want to exchange it for some new productions, is precisely equal to the quantity of labour which it can enable them to purchase or command' (Smith 1776, p. 48).

The choice of an objective standard of wealth – in Smith's case the labour 'commanded' by goods – focused the discipline of economics on the processes by which these embodiments of wealth were amassed. By the 17th century, the rise of a market organization of trade and production had already brought to the fore the distinctively 'economic' problem of wealth – namely, the need to explain its accumulation as the outcome of impersonal processes rather than as the spoils of power. From Smith's Physiocratic predecessors through John Stuart Mill, the principal aim of political economy was accordingly to investigate the consequences of a competitive struggle for wealth, with respect both to its distribution among individuals and social classes, and to its effect on the development of the system as a whole.

Almost from the outset, however, the conception of wealth as an objective element in the economic process posed troublesome questions. One of these was the appropriate treatment of labour that produced services rather than tangible goods. Because services are flows, they cannot be included in wealth, if the latter is defined as a tangible stock. The difference, as Cassel explained, involves *time*: stocks are present in their entirety at a moment in time; flows require the passage of time (Cassel 1918, p. 31). A second difficulty concerned the classification of different kinds of labour. Smith, for example, differentiated between productive and unproductive labour, calling 'productive' only the labour whose product could be sold to replenish the working capital of the manufacturer, and designating as 'unproductive' all services – 'how honourable, how useful, or how necessary soever' – because these activities consumed, but did not renew, the fund of circulating capital whence they derived their subsistence (Smith 1776, p. 331).

In addition, Smith and Ricardo recognized that labour was itself a heterogenous rather that a simple 'substance', and that some means would have to be found to reduce its complexity to a uniform basis. Both consigned the solution of the problem to the workings of the market. This may have been adequate for a rough and ready explanation of wage differentials originally established by market considerations and subsequently perpetuated by social inertia, but it concealed the deeper problem of reducing a spectrum of labour skills to a common denominator of 'simple' labour without recourse to market forces – that is, to supply and demand.

Finally, as Marx was to point out, the classical economists did not perceive that labour was a concrete activity – the labour of Ricardo's deer

hunter not being substitutable for that of his salmon fisherman – so that a level of 'abstract' labour had to be posited if labour was to serve as a universal equivalent, or measuring rod, for wealth. Although the full difficulty of reducing labour to its abstract essence escaped Marx himself, for all these reasons the concept of labour as a simple and self-evident metric became increasingly difficult to accept.

In accounting for the decline of the objective view of wealth, however, it is likely that the difficulties enumerated above did not play so important a role as another, quite separate, objection. This was the awareness that wealth as an objective entity did not express the attribute of goods that seemingly endowed them with desirability, namely their capacity to yield pleasure or utility to their possessors or beneficiaries. Oddly enough, we can also trace this view of wealth to Smith, who declared that 'Every man is rich or poor according to the degree in which he can afford the necessaries, conveniencies, and amusements of human life' (Smith 1776, p. 47).

It was Ricardo who first pointed out the inconsistency in Smith's views, in that the subjective enjoyments yielded by wealth – its 'riches' – were not the same as the expenditure of labour power required for its creation – its 'value'. Thus for Ricardo, two countries might be equally 'rich' in necessaries and conveniences, but the value of the riches of one would be larger than that of the second if they required more labour to produce (Ricardo 1821, ch. 20).

Ricardo's distinction between riches and value marks a sharp distinction between subjective (enjoyment) and objective (embodiment) conceptions of wealth, but Ricardo himself did not pursue the analytic and conceptual horizons opened up by the subjective view. That was to be the work of the post-classical period, culminating in the marginalist 'revolution' of the 1870s. Although this episode is famous for its shift in the concept of value from labour to utility, it is apparent that this shift entailed an equally deep-seated and far-reaching change in the conception of wealth, and as a consequence, in the study of economics. The works of Gossen, Menger, Jevons and Walras – the pioneers in this redirection of economics – display considerable variations in their internal details but not in their underlying depiction of the task of economics. This was now seen as an examination of the conditions for the optimization of enjoyments (utilities), not for the maximization of tangible wealth (capital). Thus Jevons wrote in *The Theory of Political Economy,* 'The problem of economics may, as it seems to me, be stated thus: Given, a certain population, with various needs and powers of production, in possession of certain lands and other sources of material: required, the mode of employing their labour which will maximize the utility of the produce' (Jevons 1871, p. 254, original in italics).

A striking consequence of this shift was the necessary divorce of economics from any quantitative estimation of the extent of wealth. Utility in the post-classical sense was not the same as the 'use-values' that had always been recognized by Smith or Ricardo or Marx as the prerequisites of exchangeability. Their use-values referred to *objective* attributes of goods – the hardness of diamonds, the softness of cloth – from which was derived the capacity of commodities to yield subjective satisfactions. The utilities of the marginalists, on the other hand, referred exclusively to the states of mind induced by the possession or use of objects. Unlike use-values, therefore, utilities were subject to continual, possibly radical shifts, induced by changes in tastes or income or the relative scarcities of objects – in all cases, changes in the relation between possessors and objects, and not changes in the physical character of the goods themselves.

From this perspective, utility therefore had no objective existence whatsoever. 'We can never say absolutely that some objects have utility and others have not', Jevons wrote; and following in that line, Robbins declared in *The Nature and Significance of Economic Science* (1932, p. 47) that 'wealth is not wealth because of its substantial qualities. It is wealth because it is scarce'.

The emphasis on the psychological element of wealth and on the role of scarcity in conferring desirability to goods clarified many questions, for example the ancient water–diamonds paradox. In addition the utility approach appeared to resolve

the problem of valuation at a level of greater generality than labour. It could be used, for example, to explain exchange value in the case of goods that required little or no labour, such as Ricardo's 'rare statues and pictures', within the same analytic framework as in the case of goods in which labour constituted a major element of cost. Thus the rise of a subjective orientation to wealth and value – we can by now surely appreciate their inextricable association – seemed an immense liberation to economists who had struggled within the constraints of an objective theory of wealth and value, whether exclusively denominated in labour or not.

The new orientation was not, however, without its problems. In so far as marginal utility is normally a direct function of scarcity, its adoption as the metric of wealth entailed the awkward conclusion that wealth as a sum of enjoyments and conveniences might well *increase* as a consequence of the *diminution* of material abundance. Some of the marginalists accepted this result; others, such as Menger, called it only an 'apparent' paradox, on the grounds that the continual augmentation of goods would gradually remove them from the category of 'economic' goods, thereby excluding them from consideration as wealth (Menger 1871, 111). This seems a question-begging resolution. In addition, the replacement of an objective by a subjective standard of wealth led to the even more awkward conclusion that the aggregation of the wealth of individuals was impossible on the same grounds as the aggregation of their feelings or experiences. It was such considerations that led Robbins to declare in his influential essay mentioned above that 'in any rigid determination of Economics, the term wealth should be avoided' (Robbins 1932, p. 47n).

All attempts to define wealth have therefore led to difficulties and even paradoxes. The conceptual and mensurational problems of an objective approach denominated in labour have been equalled, perhaps even surpassed, by those of a subjective approach denominated in utilities. Notwithstanding Robbins's reservations, however, economists have not abandoned the use of wealth as a fundamental constitutive element of economics, nor have they given up attempts to measure it. Here we can trace the general line of development once more to Adam Smith, this time to his famous abandonment of the category of labour as the measuring rod of value and his substitution of a cost of production measure which simply added up the income flows – wages, rents and profits – accruing to the three major classes.

From Ricardo on, Smith has been accused of circularity or inconsistency in this choice of an 'adding-up' approach to value, in which no attempt was made to discover a common denominator of wealth. But as a practical solution to the problem of measuring a concept that was universally regarded as real and important, whatever its intrinsic difficulties, Smith's approach was not without merit. The cost of production, or adding-up, approach to national wealth provided a common sense basis for the representation of national power or collective well-being, regardless of the unexamined problems behind these representations.

At all events, in modern times the measurement of wealth has become a major preoccupation for virtually all advanced nations. In *The Statistical Abstract of the United States,* for example, we find time series of various stocks and flows that have been selected as being of particular significance for the measurement of national wealth. The stocks include such items as estimates of financial and real assets, business and residential capital, consumers' stocks of durables, land and selected government assets; while the flows concentrate on gross national product and its components. These items have been selected partly on the basis of the availability of data and partly on the basis of their importance for national economic policy. They are neither a complete nor a consistent set of accounts, a number of important stocks and flows being absent, such as the stock of human capital, or the flow of unpaid labour in household work. The method of valuing both stocks and flows also differs from the public to the private sector, since the standard valuation is that of 'market values', which cannot apply to public goods or services. This is not the place to discuss the problems of national accounting, but it is worth noting

that the same standard of practicality is applied as we find in Adam Smith, as well as the same absence of any firm conceptual foundation.

There remains another aspect to the concept of wealth. It is expressed with his usual vigour by John Bates Clark in *The Distribution of Wealth*: 'Amounts of wealth are usually stated in money . . . The thought in the minds of men who use money as a standard of value runs forward to the power that resides in the coins. The intuitions that are at the basis of this popular mode of speech are nearer to absolute truth than much of economic analysis. They discern a power of things over men . . .' (1908, p. 376).

The aspect of wealth to which Clark directs attention is once more anticipated by Smith, who writes, 'Wealth, as Mr. Hobbes says, is power . . . the power of purchasing; a certain command over all the labour, or over all the produce of labour which is then in the market' (Smith 1776, p. 48). This definition contains an insight of great significance. To the extent that wealth is a form of power, its inadequate denomination in terms of labour commanded or utilities generated becomes explicable by virtue of the inapplicability of either metric to the 'substance' in which power must be measured.

What might that substance be? Smith and other early investigators of the nature of wealth-seeking society assumed it to be the expression of a universal desire to be admired. ' The rich man glories in his riches, because he feels that they naturally draw upon him the attention of the world ... and he is fonder of his wealth, upon this account, than for all the other advantages it procures him', Smith wrote in *The Theory of Moral Sentiments* (1759, pp. 50–1).

What was unknown to Smith, or to others, like Senior, who followed his general lead in the psychology of wealth, is that prestige and wealth do not seem to be universally conjoined. Contemporary anthropologists emphasize that wealth differs in a crucial respect from prestige in that the defining characteristic of wealth is its ability to confer social power on its possessors, whereas the enjoyment of prestige carries no such intrinsic rights. As a consequence, we find that in primitive societies, where there is universal access to the resources needed for subsistence, wealth does not exist as a social category, in that no individual or group enjoys command over the labour or the product of others, save for the claims conferred by relations of kinship or communal obligation (Sahlins 1972; Fried 1967).

From the anthropological viewpoint, then, primitive societies enjoy Ricardian riches, but no Smithian or Marxian value. From this vantage point, 'wealth' ceases to appear as an eternal attribute of human society, whether as tangible goods or the utilities enjoyed by their beneficiaries. Rather, the crucial element in the conception of wealth, and in the constitution of economics as its study, lies in the historical advent of the institution of property, construed as the right to exclude others from the material or other resources to which legal title has been gained. From this perspective, the fundamental problem posed by wealth is that of tracing the evolution of the social stratification characteristic of all post-primitive societies. Wealth is the economic face of that political stratification, lodged in the hands of a class whose ability to grant or deny access to resources becomes the 'economic' basis for both prestige and power.

See Also

- ▶ Capitalism
- ▶ Socialism

Bibliography

Cassel, G. 1918. *Theory of social economy.* New York: Harcourt Brace & Co., 1932.

Clark, J.B. 1899. *The distribution of wealth.* Boston: Ginn & Company, 1908.

Fried, M. 1967. *The evolution of political society.* New York: Random House.

Jevons, W.S. 1871. *The theory of political economy,* ed. R.D. Collison Black. Harmondsworth: Penguin, 1970.

Menger, C. 1871. *Principles of economics.* New York: New York University Press, 1981.

Ricardo, D.1821. *On the principles of political economy and taxation*, ed. P. Sraffa. Cambridge: Cambridge University Press, 1951.

Robbins, L. 1932. *An essay on the nature of significance of economic science*. London: Macmillan.

Sahlins, M. 1972. *Stone age economics*. Chicago: Aldine.
Smith, A. 1759. *The theory of moral sentiments,* ed. D.D. Raphael and A.L. Macfie. Oxford: Clarendon Press, 1976.
Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations,* ed. R.H. Campbell, A.S. Skinner and W.B. Todd. Oxford: Clarendon Press, 1976.

Wealth Constraint

Michael Allingham

Consumer theory, and thus general equilibrium theory, sees agents as making transactions subject to a wealth constraint, that is, subject to the value of their purchases not exceeding that of their sales.

The wealth constraint first appears explicitly in the writings of the neoclassical school, and particularly in the work of Walras (1874–7). However, it is first made use of rigorously by Slutsky (1915) and then Hicks (1946). In the present discussion we first specify the constraint in the pure exchange case, where borrowing and lending are only implicit, then introduce time, and finally examine the constraint in a monetary economy, where borrowing and lending become explicit.

An agent in an economy with n commodities is represented by his preferences for and endowment of these commodities. His preferences are represented by a complete preordering on the commodity odity space R^n_+, and his endowment by a point e in $S = R^n_+ - 0$. At any given price system p in S the agent chooses a most preferred point $x(p)$ in R^n_+ subject to the value of this not exceeding the value of his endowment, that is $p{\cdot}e(p) \leqq p{\cdot}e$ This constraint, which we may also write as $p{\cdot}g(p) \leqq 0$ where $g(p) = x(p) - e$ is excess demand, is his wealth constraint.

Two aspects of this require comment. The first is that the constraint is not affected if prices are multiplied by any positive scalar. The second is that, provided that all commodities are desirable, the constraint will always be binding. The relevance of these observations is apparent if we aggregate excess demands of individual agents to obtain aggregate excess demand, $f(p)$. The first aspect implies that f is homogeneous, that is, $f(tp) = f(p)$ for all positive t, while the second implies that f obeys Walras's Law, that is $p{\cdot}f(p) = 0$ for all p.

The interpretation of the budget constraint, that the agent cannot spend more than his wealth, raises the question of his borrowing (or lending). This question may be avoided by defining commodities in such a way that they are distinguished by date, that is, the time at which they are available, as well as on the basis of their physical nature. Thus not only is wheat distinguished from oats, but wheat available today is also distinguished from wheat available tomorrow. In this framework the agent borrows by selling wheat today for delivery tomorrow, and lends by buying wheat today for delivery tomorrow.

A further extension of the way in which commodities are defined can deal with the problem that future endowments may be uncertain. Under this extension commodities are also distinguished by the state of nature in which they are available. Such a state is a complete specification of everything that may affect prices, being defined in such a way that exactly one state must occur. Then if an agent will have as part of his endowment one unit of wheat tomorrow if it is not raining, and nothing if it rains, his wealth is simply the price (today) of wheat tomorrow if it is not raining. Thus although his future endowments are uncertain his wealth constraint is certain (Debreu 1959).

However, this depends on the existence of suitable contingent markets, and for reasons such as differential information, and indeed transaction costs, such markets may not exist. If they do not then the agent's intertemporal wealth constraint is uncertain: or in other words, he faces two separate constraints – one for today and one for tomorrow.

A similar problem arises if capital markets are imperfect, that is, if there is no market for a machine, say, that produces goods (even with certainty) both today and tomorrow. Again then, the agent faces separate constraints for each period. Equivalently, we may say that he has a sequence of income constraints rather than one wealth constraint.

Lying behind these problems are, implicitly, restrictions on borrowing and lending. These activities become explicit if we introduce money into the economy. Exchange without money may be cumbersome, and involve an agent who plans to exchange oats for wheat finding one agent who both wants to acquire oats and surrender wheat. With money, however, he need only find some agent who plans to buy oats (that is, exchange money for oats) and some, usually different, agent who plans to sell wheat (that is, exchange wheat for money).

The essential aspect of money (or rather economies in which all exchanges must be for money) follows from this: an agent who plans to exchange oats for wheat must sell his oats before he can buy his wheat (unless he has a sufficient reserve of money), and if he cannot sell his oats then he cannot buy wheat. Thus planned demands only become effective if they are backed by money (Clower 1967).

As is well known, this may lead to effective excess demands for all commodities being zero at the same time as planned excess demands are substantial, that is, to a state of chronic excess demand or supply (such as unemployment). However, this will not occur if borrowing and lending can occur freely, when the wealth constraint again becomes the only restriction on preferences.

See Also

► General Equilibrium

Bibliography

Clower, R.W. 1967. A reconsideration of the microfoundations of monetary theory. *Western Economic Journal* 6: 1–8.

Debreu, G. 1959. *Theory of value*. New York: Wiley.

Hicks, J.R. 1946. *Value and capital*, 2nd ed. Oxford: Clarendon Press.

Slutsky, E. 1915. Sulla teoria del bilancio del consumatore. *Giornale degli Economisti* 51: 1–26.

Walras, L. 1874–7. *Eléments d'économie politique pure*. Definitive edn. Lausanne, 1926. Trans. W. Jaffé as *Elements of pure economics*. London: George Allen & Unwin, 1954.

Wealth Effect

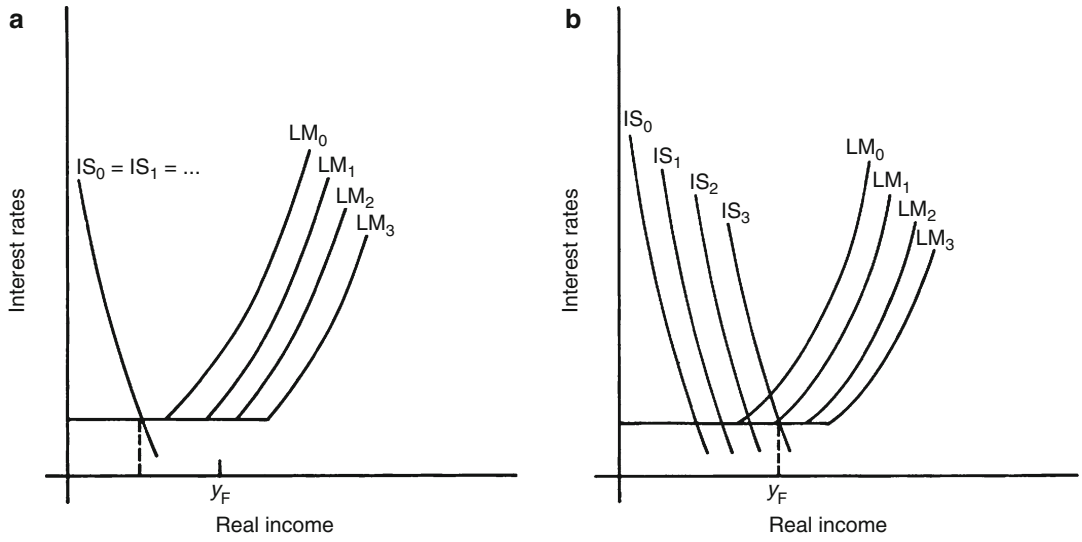
Michael R. Darby

Wealth is ubiquitous in economic analysis; so it is not surprising that the term 'the wealth effect' has been used by different authors and even the same author to refer to many distinct concepts. This brief entry is concerned with only the wealth effect in terms of aggregate consumption. Even in this limited field several distinct usages must be considered.

Haberler (1939), Pigou (1943), and Patinkin (1956) pioneered the idea that variations in money balances will cause movements, *ceteris paribus*, in aggregate consumer expenditures. This wealth effect is often referred to as the Pigou effect or real balance effect. If such a positive relationship between consumption and real money balances were to exist, these authors argued, it would profoundly affect the Keynesian analysis of underemployment equilibrium.

Recall that an underemployment equilibrium was supposed to occur if, in Hicksian terms, the IS curve intersected the LM curve in the liquidity-trap region at an income y less than fullemployment income y_F. This is illustrated in Fig. 1a where the initial LM curve is LM_0. If prices fall because of the depressed conditions, this increases real money balances, and shifts the LM curve out to LM_1, LM_2 …, but the interest rate remains fixed at the minimum rate r_{min} so that lower prices do not alleviate the depression.

Now, suppose that higher real money balances increase consumption, *ceteris paribus*. If the price level falls and real money balances rise, the IS curve shifts out to IS_1, IS_2, …. As illustrated in Fig. 1b, this progressively increases real income until IS_3 and LM_3 intersect to determine $y_3 = y_F$. Thus there are market forces which tend to restore real income to full employment levels. Other, stronger forces eliminating underemployment have been recognized subsequently, but the real balance effect was the first effective challenge to the Keynesian structure.



Wealth Effect, Fig. 1 (**a**) No wealth effect, (**b**) With wealth effect

The source of the wealth effect is a more controversial issue than its implications. The modern theory of the consumption function views wealth as the principal determinant of aggregate consumption. Since money balances are a component of wealth, it appears that a change in real money balances should have the same effect on consumption as an increase in wealth reflecting, say, a higher capital stock. Uncertainties arise, however, in regard to what portion of money is an element of wealth and the extent, if any, that the implicit service stream from money should be deemed consumption analogously to the service stream from a consumer's automobile. In addition, a portfolio adjustment element may play a significant role in the real balance effect in disequilibrium.

The first set of issues concerns whether only outside money (commodity of non-interestbearing government money) should be included in wealth. The remaining inside money is offset on banks' books by debts which increase in real value at the same time a fall in prices or increase in nominal quantity increases the real value of inside money. Accordingly, for society as a whole increases in inside money are not a source of increases in wealth, and the wealth effect is confined to outside money.

A possible measure of wealth is outside money plus physical capital plus a fraction λ of government debt where λ varies from 1 to 0 according to the extent that individuals see through the veil of government. If government bonds are not net wealth, an open market purchase of bonds with new outside money increases wealth so that both the IS and LM curves shift to the right. Thus monetary policy is strengthened by the wealth effect. However, the effect of a lower price level in shifting the IS curve to the right is strongest if λ is close to 1 since then the higher real values of both money and government bonds increases consumer spending. Evidence that λ lies near or at 0 thus favours the effectiveness of monetary policy but reduces the probable significance of price variations for aggregate consumption.

Money – particularly outside money – is very much like a durable good: its primary yield is in the form of a service stream. To the extent that outside money is held by consumers, this stream may be counted as a form of consumption. Accordingly, the increase in aggregate consumption associated with an increase in real outside money may be somewhat less than that associated with an increase in the capital stock which is held for its pecuniary income stream.

A final difficulty with interpreting the real balance effect is that higher real money balances may increase consumer expenditures by a much greater amount than pure consumption via a portfolio adjustment channel in disequilibrium. Since purchases of consumers' durable goods can be considered saving and investment by households, accelerating (retarding) these purchases may be a convenient way of reducing (increasing) cash balances greater (less) than their long-run desired value.

This disequilibrium view was suggested by Friedman and Schwartz (1963). It rationalizes both an effect for inside as well as outside money and also coefficients on real money which are large relative to any reasonable real yield on wealth (see, for example, Darby 1977–8).

In conclusions, the wealth effect on aggregate consumption has concerned macroeconomists for half a century. Unresolved questions reflect the unsettled issues of the consumption function: Is consumer behaviour described as reflecting Ricardian equivalence? What determines variations in consumers' purchases of durable goods as opposed to the consumption of their services? High nominal deficits in the 1980s have stimulated interest and provided data which may ultimately reduce our uncertainty about the first of these questions. This in turn may simplify the analysis of the second.

See Also

► IS–LM Analysis

Bibliography

Darby, M. 1977–8. The consumer expenditure function. *Explorations in Economic Research* 4(5): 645–74.

Friedman, M., A.J. Schwartz, and A.J. Schwartz. 1963. Money and business cycles. *Review of Economics and Statistics* 45(1): 32–64.

Haberler, G. 1939. *Prosperity and depression*, 2nd ed. Geneva: League of Nations.

Patinkin, D. 1956. *Money, interest, and prices*. Evanston: Row, Peterson & Co.

Pigou, A.C. 1943. The classical stationary state. *Economic Journal* 53: 343–351.

Webb, Beatrice (1858–1943) and Sidney (1859–1947)

J. M. Winter

Beatrice and Sidney Webb were the primary authors of the political economy of Fabian socialism. Their work together spanning half a century left an indelible mark on the evolution of opinion and legislation on trade unionism, local government, the relief of poverty, and theories of the transition to socialism.

The appropriate context in which to place the work of the Webbs is in the intellectual history of the labour movement. Judged by European standard, the role of intellectuals in the British labour movement has not been particularly distinguished. There are no figures in this country's past to match the stature of Jaurés, Kautsky or Gramsci, and as is well known, the debate over revisionism largely (but not completely) bypassed English labour. Partly because of their own social distance from organized labour, and partly because of *ouvrieriste* suspicions of their middle-class manners, labour's intellectuals were more isolated and politically marginalized in 19th- and 20th-century Britian than in any other European country.

The Webbs succeeded only in part in overcoming these obstacles. They chose two characteristic forms for their work: scholarship and 'permeation', by which they meant the slow and subtle conversion of men of power to their views.

First, their scholarship. One of the Webbs' major achievements was to provide the British labour movement with a record and a sense of its history. One of the primary challenges which have faced the leaders and rank-and-file of organized labour in Britian all too regularly in this century has been how to deal with political failure. What the Webbs and other historians of labour, such as G.D.H. Cole and R.H. Tawney, have offered is a way of coping with political despair, a means of surviving reversals by locating them within a long and complex experience, the parameters and

outcomes of which are always bound to remain obscure to contemporaries. The retrieval of lost strikes, abortive insurrections, and collapsed Utopian experiments can give heart to today's activists by showing them how other men and women in the past faced similar defeats and still managed to retain their convictions. Labour history, largely pioneered by the Webbs, was (and remains) an assertion of the dignity of defiance.

The academic study of the subject of industrial relations was largely begun by the Webbs. Their *History of Trade Unionism* appeared in 1894 and was reissued in four subsequent extended and revised editions in 1896, 1902, 1911, and 1920. This book was followed in 1897 by *Industrial Democracy*, which the authors claimed was not history at all, but rather an account of the structure, function and theory of contemporary trade-union organization. The nine mammoth tomes of the *History of English Local Government* appeared between 1906 and 1929. In this monumental work, they traced the evolution of parochial and county administration from the medieval period to 1929, when authority over public assistance was finally transferred from boards of guardians to county and county borough councils. In a sense, this history is a circular grand tour of the evolution of collective provision for the poor: from parish control in the medieval period, to the reduction of local government to the barren application of the principle of less eligibility in the 19th century, to the 20th-century return of local authorities to their prior position as integrated agencies of social provision.

The two sides of their historical work reflect clearly aspects of their socialist political economy. In the Webbs' political theory, men are considered first of all as consumers, and only secondarily as producers. They never fully explored the implications of this distinction, but it is at the heart of their position. On the one hand, they were the chroniclers *par excellence* of workplace loyalties and organizations. But on the other hand, they argued that the first interest of all inhabitants is the provision of consumer needs. The function of government is to provide a framework within which men can obtain the necessities of daily life. Hence they defined the state as the 'national association of consumers' engaged in 'housekeeping on a national scale'.

The collective ownership of the means of production was a means to fulfill that function in such a way as to ensure that public service rather than private profit governed the economic life of the nation. Thus their work in socialist politics was an outgrowth of their political economy, adumbrated in pamphlets and articles for the Fabian Society, which Sidney helped found in 1884, and the Labour party, the reconstruction of which in 1918, with a commitment to nationalization as a goal, was largely Sidney's major political achievement. He served as President of the Board of Trade in the 1924 Labour government and as Secretary for the Colonies and Dominions from 1929 to 1931, but without distinction in either role.

After the onset of the world economic crisis, the Webbs turned away from many of their earlier commitments and followed a disastrously misguided inclination to believe that Stalin had inaugurated a 'new civilization' in the Soviet Union in the 1930s. But even this misperception – in which the Webbs were not by any means alone – did not obscure the range and depth of the influence of their previous work. This was primarily in the field of political theory and history, rather than in economic theory, the characteristics of large tracts of which they remained blissfully ignorant for most of their lives. It was not for any incisive critique of Marx or Marshall that the Webbs will be remembered, but rather for a lifetime of service to Fabian socialism.

The legacy of their ideas, their writing and their administrative and political efforts was subtle and indirect. But, in essence, it was to provide an intellectual justification for the transition to socialism in Britain, rooted not in class struggle but in the slow, but in their view inexorable, reconstruction of the state as the servant of the basic needs of the community as a whole.

Aside from Clause Four of the Labour party constitution, the most enduring monument to the Webbs' influence is the London School of Economics. Founded in 1895, as Beatrice noted, 'on the lines of the Ecole Libre des Sciences Politique in Paris', the School was devoted not to the study of

socialism, but rather (in the words of its first prospectus) to further 'the study and investigation of the concrete facts of industrial life and the actual working of economic and political institutions as they exist or have existed, in the United Kingdom or in foreign countries'. Such impartial and scientific study of social questions would, Sidney believed, inevitably further the cause of socialism; hence there was no need to insist upon doctrinal purity among the academic staff of the School. In fact, the first director, W.A.S. Hewins, was an economist who shared none of the Webbs' assumptions, and in subsequent years, this tradition of political heterogeneity among the directors and lecturers of the School has never been challenged.

It is perhaps best to see this aspect of the Webbs' work – 'the biggest single enterprise of Our Partnership', as Beatrice put it – as reflecting two themes. The first is their lifelong interest in questions of educational provision, a subject to which Sidney devoted considerable attention during his period on the London County Council in the 1890s. The second is as part of their Fabian commitment both to the systematic analysis of contemporary social problems and to the creation of the administrative class which, they believed for most of their lives, would lead Britain out of the morass of industrial capitalism.

See Also

► Fabian Economics

Selected Works

1894. *History of trade unionism*. London: Longmans & Co.
1897. *Industrial democracy*. London: Longmans & Co.
1898. *Problems of modern industry*. London: Longmans & Co.
1906–29. *English local government from the revolution to the* M*unicipal corporations act*, 8 vols. London: Longmans & Co.
1920. A constitution for the socialist commonwealth of Great Britian. London: Longmans & Co.
1923. *The decay of capitalist civilisation*. London: Allen & Unwin.
1935. *Soviet communism. A new civilisation?*, 2 vols. London: Longmans.
1978. *The letters of Sidney and Beatrice Webb*, 3 vols. ed. N. Mackenzie. Cambridge: Cambridge University Press for the London School of Economics.
1982–4. *The diary of Beatrice Webb*, 3 vols, ed. N. and J. Mackenzie. London: Virago in association with the London School of Economics and Political Science.

Bibliography

Cole, M.I. (ed.). 1949. *The Webbs and their work*. London: Frederick Muller.

Weber, Alfred (1868–1958)

Martin Beckmann

Keywords
Launhardt, C. F. W.; Location theory; Transport costs; Weber, A.; Weber, M.

JEL Classifications
B31

German economist and cultural sociologist, born 30 July 1868 in Erfurt; died 2 May 1958 in Heidelberg. He studied in Berlin and was Privatdozent there 1900–1904, Professor in Prague 1904–7 and in Heidelberg from 1907 until his death, interrupted by voluntary retirement between 1933 and 1945.

Weber's *Reine Theorie des Standorts* (1909) established him as the leading location theorist since Von Thünen, although his theoretical model has been anticipated by Wilhelm Launhardt (1882). Weber's interest in the location of the emerging modern industries arose from his

earlier study of home industries, their struggle for survival, and the resulting social conditions of those working under the putting out system. Only Part I was published, an intended empirical Part II never appeared. Under Alfred Weber's guidance several theses were written on particular industries. Alfred Weber's article in the *Grundriss der Sozialoekonomik,* Part VI (1922) is a restatement of his book.

After the First World War, Weber turned to 'cultural sociology as cultural history' (1935). This work, while overshadowed by that of his more famous brother, Max (1864–1920), and lacking its precision, provides a fresh perspective of the development of Western civilization.

Alfred Weber, although not an impressive speaker, was a highly influential teacher. Together with Karl Jaspers he re-established academic traditions of excellence at the University of Heidelberg in the post-Second World War years.

The location of industries according to Weber is governed by cost minimization. When production costs are independent of location, this means the minimization of transportation costs. In the case of two resource deposits and one market, the optimal location may fall inside the triangle spanned by the three given locations. Economies of joint location, based on the exchange of intermediate goods or the joint use of indivisible facilities induce 'agglomeration' of industrial activities in large centres.

See Also

- ▶ Location Theory
- ▶ Thünen, Johann Heinrich von (1783–1850)

Selected Works

1909. *Reine Theorie des Standorts.* Trans. C.J. Friedrichs as *Theory of the location of industries*. Chicago: University of Chicago Press, 1929.

1935. *Kulturgeschichte als Kultursoziologie*, 2nd ed. Munich: Piper, 1950.

Bibliography

Salin, E. 1958. Obituary. *Kyklos* 11: 318–340.

Weber, Max (1864–1920)

David Beetham

Abstract

Max Weber was one of the leading historical political economists in the Germany of the 1890s. Weber's early work in political economy reflected the distinctive concerns of a younger generation of the Historical School, which sought to demonstrate the theoretical character of the concepts used in historical economics and the historical presuppositions of theory. Over time, Weber's research became increasingly wideranging and theoretical, involving an elucidation of the character of Western rationalism as applied to the basic structures of economy and society, and reflecting a shift in his disciplinary focus of interest from political economy to sociology.

Keywords

Bureaucracy; Capital accumulation; Class; Class conflict; Conspicuous consumption; Division of labour; Exploitation; Factor demand; Historical materialism; Historical School; German; Individualism; Labour theory of value; Managerial revolution; Menger, C; Modernization; Planning; Private property; Protestant ethic thesis; Surplus value; Value judgements; Weber, M.

JEL Classifications

B31

Max Weber was born in 1864 at Erfurt and died in 1920 at Munich. After early studies in the history of commercial law, he established himself as one of the leading figures in a new generation of

historical political economists in the Germany of the 1890s. He was appointed to chairs in political economy at Freiburg in 1894 and at Heidelberg in 1896. A nervous breakdown commencing in 1898 led to his withdrawal from academic teaching, but did little to impair the flow of his writing. In 1904 he took over the editorship of the *Archiv für Sozialwissenschaft und Sozialpolitik*, the leading academic journal in 'social economics', devoted to the exploration of the interrelationship between economy on the one hand, and law, politics and culture on the other. This interconnection formed the main site of Weber's own research, whose focus became increasingly wide-ranging and theoretical, involving an elucidation of the character and presuppositions of modern Western rationalism, as applied to the basic structures of economy and society. Weber was also actively and often controversially involved in the political issues of Wilhelmine Germany, from a progressive national-liberal standpoint, and during the war was one of the leading polemicists for a democratization of the constitution. Such involvement gave particular sharpness to his discussions of social science methodology, the role of value judgements and the relation between academic analysis and political practice. It was only comparatively late in his life that he came to think of his work as 'sociology', though it is as one of the 'founding fathers' of sociology that he is now known. He resumed full-time teaching activity as professor at Munich in 1919 only shortly before his death.

Weber's early work in political economy can best be understood as reflecting the distinctive concerns of a younger generation of the Historical School (including Schulze-Gävernitz, Sombart, Max and Alfred Weber). At the methodological level they sought to resolve the controversy between the theoretical and historical schools by demonstrating the theoretical character of the concepts used in historical economics on the one hand, and the historical presuppositions of theory on the other. An important element in this resolution was to secure the acceptability of the Marxian concept of 'capitalism' as a valid concept for economic analysis, despite the untenability (as they saw it) of the labour theory of value, and exaggerated claims made for the materialist conception of history.

The recognition of the conflict between labour and capital as a systemic property of capitalism was central to Weber's early work. In his study of the impact of capitalist organization on the agricultural estates east of the Elbe, it supported his conclusion that class conflict had permanently undermined the economic basis of Junker political dominance in the Reich (Weber 1892). From it he also derived a position on social policy which was critical of the paternalism of the 'Kathedersozialisten', arguing instead that trades unions should be given a secure legal status so that they could bargain for themselves on a more equal footing with capital. The distinctive Weberian conception of class conflict under capitalism was theorized neither in the Marxist terms of 'exploitation' nor in the neoclassical terms of 'factor demand', but in terms of a systematic competition for the social product on the basis of a power relation between the classes that was adjusted and underwritten at the political level.

If the incorporation of Marxian insights into the mainstream of social economics required that the analysis of class conflict be freed from the doctrine of surplus value, it also required a critique of the one-sided assumptions of historical materialism. This Weber offered in his most famous study The *Protestant Ethic and the Spirit of Capitalism* (1904). The argument of this work was that the profit-maximizing behaviour so characteristic of the bourgeoisie, which could be explained under fully developed capitalist conditions by its sheer necessity to survival in the face of competition, could not be so explained under the earlier phases of capitalist development. It was the product of an autonomous impulse to accumulate far beyond the needs of personal consumption, an impulse which was historically unique. Weber traced its source to the 'worldly asceticism' of reformed Christianity, with its twin imperatives to methodical work as the chief duty of life, and to the limited enjoyment of its product. The unintended consequence of this ethic, which was enforced by the social and psychological pressures on the believer to prove (but not earn) his

salvation, was the accumulation of capital for investment.

Early critics of Weber's thesis misunderstood it as a purely cultural explanation for capitalism, as if 'a Siberian Baptist or a Calvinist inhabitant of the Sahara' must inevitably become a successful entrepreneur. Weber was in fact well aware both of the material preconditions for capitalist development, and of the social interests that are needed to support the dissemination of new ideas. The crucial question about his thesis is whether the employment of wage labour that made unlimited accumulation possible in principle, also made it inevitable in practice; whether, that is, the Protestant ethic should be seen as providing a necessary motivation for capitalist accumulation, or rather a legitimation for it in the face of prevalent values favouring conspicuous consumption on the part of a leisured class. Weber himself saw it as both. His work is the most sophisticated in a long tradition of exploration of the cultural preconditions for capitalist accumulation, from Adam Smith's celebration of 'parsimony', to recent explanations of Britain's economic decline in terms of the gentrification of its entrepreneurial spirit.

At one level the 'Protestant ethic thesis' can thus be read as a critique of historical materialism, with its explanation of capitalist development as the necessary outcome of the feudal order, rather than as the result of a unique conjunction of favourable historical conditions, cultural and political as well as economic. At another level it can be read as an extended critique of the ahistorical theorizing of Carl Menger and the Austrian school. In Weber's view the methodical, calculating, welfare-maximizing behaviour of the neoclassical models was not a universal characteristic of human rationality as such, but a product of modern Western rationalism. His subsequent studies of the economic ethic of the major world religions (Confucianism, Hinduism, Buddhism, ancient Judaism; collected in Weber 1921) were designed to elucidate this distinctive cultural complex. They showed that, while instrumental rationality was a universal category of social action, only in the modern West had the goal-maximizing calculation of the most efficient means to given ends become generalized. And while other cultures had attempted to make the world intelligible through the development of elaborate theodicies, or to create internally consistent systems of ethics or law, the distinctive features of Western rationalism were the scientific assumption that all things could be comprehended by rationalism. His subsequent studies of the economic ethic of the major world religions (Confucianism, Hinduism, Buddhism, ancient Judaism; collected in Weber 1921) were designed to elucidate this distinctive cultural complex. They showed that, while instrumental rationality was a universal category of social action, only in the modern West had the goal-maximizing calculation of the most efficient means to given ends become generalized. And while other cultures had attempted to make the world intelligible through the development of elaborate theodicies, or to create internally consistent systems of ethics or law, the distinctive features of Western rationalism were the scientific assumption that all things could be comprehended by reason, together with the attitude of practical mastery which sought to subject the world to human control rather than merely adjust oneself to it. In Weber's major unfinished theoretical work, *Economy and Society* (1922), capitalism was shown to be simply one expression, rather than the unique locus, of this 'rationalization' process. The work is structured around the antithesis between 'traditional' and 'rationalized' forms of action and organization in all spheres of social life, and the transition between the two provides the key to the Weberian theory of modernization.

The conclusion of Weber's mature work, that capitalism was to be understood as part of a wider 'rationalization' process, coincided with his analysis of its most advanced forms in contemporary Germany. According to this analysis, the distinctive feature of capitalist concentration was the change in its internal mode of organization: the adoption of a complex technical division of labour and a hierarchical structure of administration that increasingly resembled the bureaucratic type already established in the political sphere. For Weber, the bureaucratic model of administration was becoming generalized throughout all sectors of contemporary society, because of its efficiency

in performing complex organizational tasks. Along with it went the emergence of a new middle class, whose distinctive position in the class stucture depended neither on property ownership (capital) nor its absence (labour), but on the possession of technical and organizational skills, and on its authority position within a bureaucratic hierarchy.

Some commentators have seen Weber as an early forerunner of the 'managerial revolution' thesis. Certainly he was among the earliest to identify technical knowledge and organization as crucial sources of social power in modern societies. But to Weber no manager could be a substitute for the risk-taking entrepreneur who stood at the head of large capitalist organizations. The bureaucratic system, with its secure career and promotion prospects, represented a conservative principle in social life, in contrast to the dynamic principle of the market. Like his junior colleague, Joseph Schumpeter, Weber saw the major source of economic innovation to be provided by the captains of industry, ready to chance their judgement in the competition of the market place. This was directly paralleled in the political sphere by his theory of competitive leadership democracy, according to which the leaders of mass parties with their bureaucratic machines competed for support in the electoral market place. If the creative force of individualism, deriving from the Protestant ethic, had itself unintentionally produced the age of organization, in which competition at the individual level was eliminated, nevertheless the role of individualism was reasserted in Weberian social theory at the head of organizations, in the form of 'charismatic' leadership.

Weber's theory of bureaucracy also provided the basis for a thoroughgoing critique of socialist planning, as prefigured in the wartime German economy. Weber was quick to echo von Mises's contention that a coherent system of allocation was impossible without market indicators, since it confirmed his own historical analysis of the preconditions for rational economic calculation. However, his distinctive critcisms of socialist planning derived from the massive extension of bureaucratic coordination he believed it would entail. Without market competition, he argued, the economy would simply stagnate. Yet the workers would remain subordinate to the same hierarchy of authority at the work place, since this was determined by the technical requirements of production, not by the particular system of ownership. Indeed, their subordination would become a new form of slavery, since the separate hierarchies of the economic, legal and political spheres would be fused into a single, all-embracing structure of power. It was the dictatorship of the official, he concluded, not of the worker, that was on the march (Weber 1918).

Overall, the progressive widening in the focus of Weber's theoretical concerns, from the conditions for economic rationality to the general theme of 'rationalization', and the subsumption of capitalism itself into the wider category of bureaucratic organization, reflected a shift in his disciplinary focus of interest from political economy to sociology. This was not just a personal development of Weber's, but one typical of the period in which he lived. With the narrowing of theoretical focus represented by neoclassical economics, it was left to the nascent discipline of sociology to take over some of the wider concerns of political economy. The rich tradition of the German Historical School, and the methodological debates which it had aroused, made German sociology particularly well placed for this enterprise. It was also particularly urgent in a country where the claims of Marxism to provide a convincing overall theory of society were widely accepted within the labour movement. It was no accident, therefore, that the most sustained rebuttal of these claims should come from the same context. As suggested above, however, Weber's approach to Marxism was not one of outright rejection, but of incorporating its insights into a different theoretical framework which left the validity of private property ownership intact. If the general presuppositions of liberalism had been thematized in the form of political philosophy in the 18th century, and of political economy in the 19th, they can be said to have received their distinctive 20th-century expression in the form of Weberian sociology.

See Also

- ▶ Bureaucracy
- ▶ Historical School, German

Selected Works

1892. Die Lage der Landarbeiter im ostelbischen Deutschland. In *Max Weber Gesamtausgabe*, vol. 1/3, ed. M. Riesebrodt. Tübingen: J.C.B. Mohr, 1984.
1904. *The protestant ethic and the spirit of capitalism*. London: Allen & Unwin, 1930.
1918. Socialism. In *Weber selections in translation,* ed. W.G. Runciman. Cambridge: Cambridge University Press, 1978.
1921. *Gesammelte Aufsätze zur Religionssoziologie*. Tübingen: J.C.B. Mohr.
1922. *Economy and society.* New York: Bedminster Press, 1968.

Bibliography

Camic, C., P. Gorski, and D. Trubek (eds.). 2005. *Max Weber's economy and society a critical companion*. Stanford: Stanford University Press.
Ringer, F. 2004. *Max Weber: An intellectual biography.* Chicago: University of Chicago Press, 2004.
Swatos, W., and L. Kaelbar (eds.). 2005. *The protestant ethic turns 100: Essays on the centenary of the Weber Thesis*. Boulder: Paradigm Publishing.
Swedberg, R. 2005. *The Max Weber dictionary: Key words and central concepts*. Stanford: Stanford University Press.
Turner, S. (ed.). 2000. *The Cambridge companion to Weber*. Cambridge: Cambridge University Press.

Weintraub, Sidney (1914–1983)

J. A. Kregel

Keywords

Aggregate demand and supply; Neoclassical synthesis; Price theory; Tax-based incomes policy; Weintraub, S.

JEL Classifications

B31

Weintraub was born in Brooklyn, New York, on 28 April 1914 and died in Philadelphia, Pennsylvania, on 19 June 1983. Professor at the University of Pennsylvania from 1950 until his death, Weintraub was widely known as the originator (with Henry Wallich) of tax-based incomes policy (TIP); his professional reputation is based on his early criticism (1959b) of the 'neoclassical synthesis' of Keynes's macroeconomics and Walrasian general equilibrium and his own highly original attempts to produce a microeconomics compatible with Keynes's macroeconomic theory.

A postgraduate year (1938–9) at the London School of Economics convinced him of the implications of Keynes's *General Theory* for price theory. His Ph.D. dissertation ('Monopoly and the Economic System', St Johns, 1941) and a series of articles on the formulation of demand in conditions of imperfect competition and imperfect information produced his innovative *Price Theory* (1949).

After the war, Weintraub concentrated on producing a micro-theory compatible with Keynes's theory of the endogeneous determination of the equilibrium level of output at less than full employment. His earlier work was extended to the demand for labour and the microfoundations of the aggregate demand and supply curves. Although this work, summarized in *An Approach to the Theory of Income Distribution* (1958), reached similar conclusions to the aggregate distribution theories then being worked out by Kaldor and Robinson, its inspiration was the formulation of a 'Keynesian' microeconomics rather than growth theory.

Evidence of the stability of the markup of prices over wage costs presented in *A General Theory of the Price Level* (1959a) led to a 'watchtower approach' to wage policy which preceded the widely discussed, but never applied, TIP proposal (1971a, b).

Weintraub's prolific writing and lecturing activities were complemented in 1978 by the founding and editing (with Paul Davidson) of *The Journal of Post Keynesian Economics*.

See Also

▶ Post Keynesian Economics

Selected Works

1942. The foundations of the demand curve. *American Economic Review* 32: 538–552.
1946. Monopoly pricing and unemployment. *Quarterly Journal of Economics* 61: 108–124.
1949. *Price theory.* New York: Pitman.
1956. A macroeconomic approach to the theory of wages. *American Economic Review* 46: 845–856.
1957. The micro-foundations of aggregate demand and supply. *Economic Journal* 67: 455–470.
1958. *An approach to the theory of income distribution*. Philadelphia: Chilton.
1959a. *A general theory of the price level*. Philadelphia: Chilton.
1959b. *Classical Keynesianism, monetary theory and the price level*. Philadelphia: Chilton.
1966. *A Keynesian theory of employment growth and income distribution*. Philadelphia: Chilton.
1971a. An incomes policy to stop inflation. *Lloyds Bank Review* 99: 1–12.
1971b. (With H.C. Wallich.) A tax-based incomes policy. *Journal of Economic Issues* 5(2): 1–19.
1978. *Keynes, Keynesians and Monetarists*. Philadelphia: University of Pennsylvania Press.

Welfare Costs of Business Cycles

Ayse Imrohoroglu

Abstract

The welfare cost of business cycles measures the benefits that would be obtained by individuals from eliminating all the macroeconomic instability in a given economy. In a seminal paper, Lucas (Models of business cycles. Yrjo Jahnsson Lectures. Oxford: Basil Blackwell, 1985) argued that these benefits are almost certain to be trivially small, especially when they are compared with the benefits that can be achieved with more growth for the post-war US economy.

Keywords

Business cycles; Great Depression; National Bureau of Economic Research; Risk aversion; Stabilization policy; Unemployment insurance; Welfare cost of business cycles

JEL Classifications

D4; D10

The welfare cost of business cycles measures the benefits in terms of additional consumption that would be obtained by individuals from eliminating all the macroeconomic instability in a given economy.

Hodrick and Prescott (1997) and Lucas (1977) define business cycles as recurrent fluctuations of output about trend and the co-movements among other aggregate time series. These fluctuations are typically represented as expansions and recessions in economic activity. The National Bureau of Economic Research, a private non-profit organization that is responsible for updating the business cycle chronology, defines a recession as 'a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real gross domestic product (GDP), real income, employment, industrial production, and wholesale-retail sales' (NBER 2003, p. 1). One of the prevailing views in macroeconomics is that business cycles are welfare reducing and governments should try to stabilize the economy by using fiscal or monetary policies.

In his seminal work, Lucas (1985) proposes a simple framework to think about how to compute the cost of economic instability, and challenges the paradigm that business cycles have large welfare costs. His measure of the welfare cost for the United States turns out to be trivially small, which disputes the need for developing more advanced policies that would eliminate fluctuations in the

United States. The following section examines his work and the subsequent research.

Lucas proposes that in order to understand the welfare cost of instability we need to start with the preferences of a hypothetical consumer who is faced with a sequence of consumption goods over time labelled $\{c_t\}$. The expected utility of such a sequence can be calculated by,

$$E\left\{\sum_{t=0}^{\infty}\beta^t\frac{1}{1-\sigma}\left(c_t^{1-\sigma}-1\right)\right\}$$

where $U(c_t)=\left(c_t^{1-\sigma}-1\right)/(1-\sigma)$ is the period utility function, β is the expectations operator, is the subjective discount factor and $\sigma>0$ is the coefficient of relative risk aversion. An important property of this utility function is that consumers would prefer smooth consumption streams to fluctuating ones or that they would prefer a deterministic consumption path to a risky path with the same mean.

In this construct, in order to understand how consumers may feel about economic instability, we can simply ask them to evaluate their lifetime utility under two different scenarios. In particular, suppose the consumers are asked to compare the lifetime utility of a perfectly smooth consumption path with a consumption stream that increases in good times and decreases in bad times while maintaining the same average level over time. The latter consumption stream is the one that results in the case of business cycles. Surely, consumers who care about smoothing consumption over time will rank the utility generated by such a stream lower than the one from the smooth consumption stream. In fact, the higher the value of σ, the lower the utility of a fluctuating consumption stream will be. With this in mind we can ask a second question. What would it cost to compensate all individuals in terms of extra consumption, uniform across time and different shocks, so that they will be indifferent between the smooth and the fluctuating consumption paths? This turns out to be a fairly easy calculation where the following equation provides a quantitative answer:

$$\lambda=\frac{1}{2}\sigma\mu^2$$

where λ is the compensation parameter, σ is the coefficient of relative risk aversion and μ measures the standard deviation in consumption.

Our hypothetical example can be made concrete by examining the properties of personal expenditures on consumption in a particular economy. Lucas (2003) uses US data for the period 1947–2001 and calculates the standard deviation of the log of real per capita consumption about the linear trend to be 0.032. Using this estimate, we can arrive at several measures of the cost of instability based on different assumptions on the coefficient of relative risk aversion. The amazing part of the findings is that the magnitude of these estimates range between one 20th of one per cent to one or two tenths of a per cent of consumption for risk aversion parameters between one and four. (Risk aversion coefficients in this range are considered to be consistent with many observations in an economy. However, much higher values are needed for some other observations such as the equity premium, which is discussed shortly.) For example, if $\sigma=2$ and $\mu=0.032$, then the consumption compensation that is required to make an individual indifferent between a fluctuating versus a constant consumption stream is about 0.001. For the US economy that would suggest that an annual consumption compensation as low as $28.96 per person would be sufficient to make individuals indifferent between a fluctuating and a smooth consumption stream. (Personal consumption expenditures in the United States in 2004 were $8.6 trillion. One-tenth of a per cent results in a total consumption compensation of $8.6 billion. Using the 2004 population of 297 million people results in consumption compensation per person of $28.96.) Such a welfare cost is negligible not only in an absolute sense but also when compared with other welfare cost measures. For example, Lucas (2000) calculates the welfare loss of a one per cent reduction in the growth rate of the economy to be as high as 20 per cent of consumption and the welfare cost of ten per cent inflation to be one per cent of income annually. Both of these estimates are more than an order of magnitude higher than the welfare cost of economic instability.

Lucas proposes to take the low cost findings seriously as giving a range of estimates for the size of the potential gains from developing policies that would eliminate fluctuations in the United States. Taking these results seriously is exactly what the profession did. Twenty years after Lucas's (1985) study, many economists continue to work on this subject, investigating whether the conclusions reached in his framework are valid under more complicated and sometimes more realistic frameworks.

Many of the assumptions in the original framework have been challenged. One of the main assumptions is that all agents are identical and have access to fully developed capital markets. One can easily imagine that, while the costs of instability may be low for some consumers, such as those with large savings, they may be devastating for some others, who may not have the means to insure themselves against these shocks. Several papers have investigated the welfare costs of instability for heterogeneous agents with limited access to capital markets. (Starting with Imrohoroglu 1989, papers that have introduced incomplete markets and examined the role of idiosyncratic risk include Atkeson and Phelan 1994; Gomez el at. 2001; Krusell and Smith 1999, 2002; and Krebs 2003.) Krusell and Smith (1999) examine an economy with substantial heterogeneity where individuals face idiosyncratic and aggregate risk and can smooth their consumption only through private savings. Their economy generates a wealth distribution that resembles US wealth distribution reasonably well. They investigate whether the welfare costs of cycles may be very high for some members of the society such as the unemployed even if in aggregate the costs are relatively low. Their findings indicate that while the welfare effects of eliminating cycles do differ across consumers they are extremely small for almost all consumers. Only for a very few individuals with almost zero consumption are welfare losses found to be as high as two per cent of average consumption.

Some of the papers in this area have highlighted the importance of understanding the interaction between aggregate and individual shocks in an economy. For example, how long-lasting are the effects of a bad shock? Do aggregate shocks compound the effects of individual shocks? Storesletten et al. (2001) show that, in an environment where small aggregate shocks can have a long-lasting impact on individuals' earnings, the welfare cost of business cycles can be much higher than the original estimates. (Beaudry and Pages 2001, also study idiosyncratic wage risk that worsens in recessions, and obtain high estimates. However, they do not allow for savings to help smooth consumption in the economy with fluctuations.) Atkeson and Phelan (1994), on the other hand, discuss the connection between aggregate and idiosyncratic risk, and suggest as a serious possibility that the elimination of aggregate risk does not affect individual risk at all. In their framework welfare cost estimates are close to zero. However, if the effects of a bad shock are assumed to be permanent, as in Krebs (2003), then the welfare costs of business cycles can be as high as 7.5 per cent of consumption. In such a framework, even if credit markets are perfect, individuals will not borrow to smooth the negative shocks they face since the effect of those shocks will persist for ever.

Another set of papers have introduced different preferences or have implicitly or explicitly used higher risk-aversion coefficients in examining the welfare cost of business cycles. While higher costs are obtained in some of these environments, there are questions about the soundness of using very high risk-aversion coefficients. For example, Tallarini (2000) finds much larger costs in a model with Epstein–Zin type preferences where preference parameters are chosen to be consistent with observed asset market data. However, the main factor behind this finding is the use of a high risk-aversion parameter to be consistent with asset price determination. (Similarly, Alvarez and Jermann 2004, find large welfare costs of economic instability in a framework that uses high risk aversion to match the six per cent equity premium in asset markets. See also Dolmas 1998; Obstfeld 1994.) Otrok (2001), on the other hand, suggests that in a model that allows for potential time- non-separabilities in preferences, which is calibrated to be consistent with observed fluctuations in a general equilibrium model of

business cycles, the welfare cost of business cycles turns out to be quite low.

It might also be possible to obtain a higher cost of fluctuations if there are links between economic growth and fluctuations. For example, Ramey and Ramey (1995) demonstrate a strong negative relationship between volatility and growth in a panel of 92 countries. However, in examining the welfare cost of instability, Epaulard and Pommeret (2003) find the volatility in the US economy to be too small to generate large benefits from stabilization policies even if reductions in volatility induce growth. Jones et al. (1999) demonstrate that the relationship between volatility in fundamentals and mean growth can be positive or negative. Their quantitative results indicate that the size of this effect is not large enough to generate large welfare costs of instability. Barlevy (2004a), on the other hand, proposes a set-up where eliminating fluctuations reallocates investment from periods of high investment to periods of low investment. This mechanism results in achieving higher growth rates without necessarily requiring higher investment levels. In such a framework, he finds the welfare cost of instability to be substantially higher than in the original Lucas estimates. The key to obtaining such large costs in his model is the presence of diminishing returns to investment, for which there is some, but not overwhelming, evidence.

It may be important to point out that the way Lucas, and Hodrick and Prescott have defined business cycles, namely, as fluctuations around a trend, has an important implication for the welfare cost calculations. If instead recessions were viewed as inefficient declines in output, as in the Keynesian view, and stabilization policies were seen as policies that would prevent economic activity from falling below its maximum potential, then the welfare cost measure could be higher. This is the case in DeLong, and Summers (1988) and Cohen (2000), who obtain welfare costs of stabilization of around 1.6 per cent and one per cent respectively. In their frameworks stabilization increases the average level of consumption.

It is important to stress that the estimates that have been discussed so far have been for the post-war US economy. The low cost estimate that is obtained in many of these papers is partly due to the relative stability of the US economy since the 1950s. Welfare costs of business cycles are higher in economies that are faced with larger fluctuations in consumption. Using the volatility of consumption in the United States prior to the Second World War, or the fluctuations in consumption that are observed in many developing countries, results in significantly higher welfare cost measures (see, for example, Pallage and Robe 2003). In addition, in the post-war period the US economy had a well-developed unemployment insurance system that may have helped reduce the volatility in consumption. Economies with less-developed welfare systems also yield higher welfare costs of instability. (Chatterjee and Corbae 2007, find that the potential benefit of reducing the likelihood of economic crises such as a Great Depression-style collapse of economic activity can range between 1.05 and 6.59 per cent of annual consumption. They also find that uninsured unemployment risk contributes significantly to the size of these gains.)

Although there is still some debate over the size of the welfare costs of business cycles, the weight of the evidence seem to suggest that they may not be too high for the US economy. (See also Barlevy 2004b, for a survey of the literature on the welfare cost of business cycles.)

See Also

- ▶ Growth and Cycles
- ▶ Liquidity Constraints
- ▶ Neoclassical Growth Theory
- ▶ Real Business Cycles

Bibliography

Alvarez, F., and U. Jermann. 2004. Using asset prices to measure the cost of business cycles. *Journal of Political Economy* 112: 1223–1256.

Atkeson, A., and C. Phelan. 1994. Reconsidering the cost of business cycles with incomplete markets. *NBER macroeconomic annual*. Cambridge, MA: MIT Press.

Barlevy, G. 2004a. The cost of business cycles under endogenous growth. *American Economic Review* 94: 964–990.

Barlevy, G. 2004b. The cost of business cycles and the benefits of stabilization: A survey. Working Paper No. 10926. Cambridge, MA: NBER.
Beaudry, P., and C. Pages. 2001. The cost of business cycles and the stabilization value of unemployment insurance. *European Economic Review* 45: 1545–1572.
Chatterjee, S., and D. Corbae. 2007. On the aggregate welfare cost of Great Depression unemployment. *Journal of Monetary Economics* 54: 1529–1544.
Cohen, D. 2000. A quantitative defense of stabilization policy. FEDS Working Paper No. 2000–34. Board of Governors of the Federal Reserve System.
DeLong, B., and L. Summers. 1988. How does macroeconomic policy affect output? *Brookings Papers on Economic Activity* 1988(2): 433–480.
Dolmas, J. 1998. Risk preferences and the welfare cost of business cycles. *Review of Economic Dynamics* 1: 646–676.
Epaulard, A., and A. Pommeret. 2003. Recursive utility, growth, and the welfare cost of volatility. *Review of Economic Dynamics* 6: 672–684.
Gomez, J., J. Greenwood, and S. Rebelo. 2001. Equilibrium unemployment. *Journal of Monetary Economics* 48: 109–152.
Hodrick, R., and E. Prescott. 1997. Post-war U.S. business cycles: A descriptive empirical investigation. *Journal of Money, Credit, and Banking* 29: 1–16. Originally a 1978 working paper at Carnegie-Mellon University.
Imrohoroglu, A. 1989. The cost of business cycles with indivisibilities and liquidity constraints. *Journal of Political Economy* 97: 1364–1383.
Jones, L., R. Manuelli, and E. Stacchetti. 1999. Technology (and policy) shocks in models of endogenous growth. Working Paper No. 7063. Cambridge, MA: NBER.
Krebs, T. 2003. Growth and welfare effects of business cycles in economies with idiosyncratic human capital risk. *Review of Economic Dynamics* 6: 846–868.
Krusell, P., and A. Smith. 1999. On the welfare effects of eliminating business cycles. *Review of Economic Dynamics* 2: 245–272.
Krusell, P., and A. Smith. 2002. Revisiting the welfare effects of eliminating business cycles. Working Paper, Carnegie-Mellon University.
Lucas, R. 1977. Understanding business cycles. In *Stabilization of the domestic and international economy,* ed. K. Brunner and A. Meltzer. Carnegie-Rochester conference series on public policy, vol. 5. Amsterdam: North-Holland.
Lucas, R. 1985. *Models of business cycles.* Yrjo Jahnsson Lectures. Oxford: Basil Blackwell.
Lucas, R. 2000. Inflation and welfare. *Econometrica* 68: 247–274.
Lucas, R. 2003. Macroeconomic priorities. *American Economic Review* 93: 1–14.
NBER (National Bureau of Economic Research). 2003. The NBER's recession dating procedure. Business Cycle Dating Committee, NBER, 21 October. Online. Available at http://www.nber.org/cycles/recessions.pdf. Accessed 27 Sept 2006.
Obstfeld, M. 1994. Evaluating risky consumption paths: The role of intertemporal substitutability. *European Economic Review* 38: 1471–1486.
Otrok, C. 2001. On measuring the welfare cost of business cycles. *Journal of Monetary Economics* 47: 61–92.
Pallage, S., and M. Robe. 2003. On the welfare cost of economic fluctuations in developing countries. *International Economic Review* 44: 668–677.
Ramey, G., and V. Ramey. 1995. Cross-country evidence on the link between volatility and growth. *American Economic Review* 85: 1138–1151.
Storesletten, K., C. Telmer, and A. Yaron. 2001. The welfare cost of business cycles revisited: Finite lives and cyclical variation in idiosyncratic risk. *European Economic Review* 45: 1311–1339.
Tallarini, T. 2000. Risk sensitive real business cycles. *Journal of Monetary Economics* 45: 507–532.

Welfare Economics

Allan M. Feldman

Abstract

Welfare economics attempts to define and measure the 'welfare' of society as a whole. It tries to identify which economic policies lead to optimal outcomes, and, where necessary, to choose among multiple optima. This article answers three fundamental qsts with three fundamental theorems. In a competitive economy, will an equilibrium outcome be optimal? Can any optimal outcome be achieved by a modified market mechanism? Is there a reliable way to measure social welfare, or to derive the preferences of society from the preferences of individuals? The negative answer to the third question is partly overcome by the theory of implementation.

Keywords

Arrow social welfare function; Arrow's th; Bergson, A.; Central planning; Command economy; Common good; Compensation tests; Competitive equilibrium; Condorcet, Marquis de; Consumer surplus; Corn Laws;

Decentralized socialism; Externalities; First fundamental theorem of welfare economics; Implementation; Industrial policy; Interpersonal utility comparisons; Kaldor, N.; Laissez-faire; Lange, O. R.; Lump-sum taxes; Lump-sum transfers; Majority rule; Mechanism design; Mises, L. E. von; Nash strategy; Pareto efficiency; Partial equilibrium; Pigou, A. C.; Pigouvian taxes; Plurality voting; Preference misrepresentation; Preference profiles; Production plans; Scitovsky, T.; Second fundamental theorem of welfare economics; Smith, A.; Social choice function; Socialism; Strategic behaviour; Third fundamental theorem of welfare economics; Voting; Voting cycles; Voting paradoxes; Welfare economics

JEL Classifications
D6

In 1776, the same year as the American Declaration of Independence, Adam Smith published *The Wealth of Nations.* Smith laid out an argument that is now familiar to all economics students: (*a*) the principal human motive is self-interest; (*b*) the invisible hand of competition automatically transforms the self-interest of many into the common good; (*c*) therefore, the best government policy for the growth of a nation's wealth is that policy which governs least.

Smith's arguments were at the time directed against the mercantilists, who promoted active government intervention in the economy, particularly in regard to (ill-conceived) trade policies. Since his time, his arguments have been used and reused by proponents of laissez-faire throughout the 19th and 20th centuries. Arguments of Smith and his opponents are still very much alive today: the pro-Smithians are those who place their faith in the market, who maintain that the provision of goods and services in society ought to be done, by and large, by private buyers and sellers acting in competition with each other. One can see the spirit of Adam Smith in economic policies involving deregulation, tax reduction, denationalizing industries, and reduction in government growth in Western countries; and in the deliberate restoration of private markets in China, the former Soviet Union and other eastern European countries. The anti-Smithians are also still alive and well; mercantilists are now called industrial policy advocates, and there are intellectuals and policymakers who believe that: (*a*) economic planning is superior to laissez-faire; (*b*) markets are often monopolized in the absence of government intervention, crippling the invisible hand of competition; (*c*) even if markets are competitive, the existence of external effects, public goods, information asymmetries and other market failures ensure that laissez-faire will not bring about the common good; (*d*) and in any case, laissez-faire may produce an intolerable degree of inequality.

The branch of economics called welfare economics is an outgrowth of the fundamental debate that can be traced back to Adam Smith, if not before. It is the economic theory of measuring and promoting social welfare.

This entry is largely organized around three propositions. The first answers this qst: in an economy with competitive buyers and sellers, will the outcome be for the common good? The second addresses the issue of distributional equity, and answers this qst: in an economy where distributional decisions are made by an enlightened sovereign, can the common good be achieved by a slightly modified market mechanism, or must the market be abandoned? The third focuses on the general issue of defining social welfare, or the common good, whether via the market, via a centralized political process, or via a voting process. It answers this qst: does there exist a reliable way to derive the true interests of society, regarding, for example, alternative distributions of income or wealth, from the preferences of individuals?

This entry focuses on theoretical welfare economics. There are related topics in practical welfare economics which are only mentioned here. A reader interested in the practical problems of evaluating policy alternatives can refer to entries on CONSUMER SURPLUS, COST–BENEFIT ANALYSIS and compensation principle, to name a few.

The First Fundamental Theorem, or, Laissez-Faire Leads to the Common Good

> The greatest meliorator of the world is selfish, huckstering trade.
> (R.W. Emerson, *Work and Days*)

In The *Wealth of Nations*, Book IV, Smith wrote: 'Every individual necessarily labours to render the annual revenue of the society as great as he can. He generally indeed neither intends to promote the public interest, nor knows how much he is promoting it. . . . He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention.' The first fundamental theorem of welfare economics can be traced back to these words of Smith. Like much of modern economic theory, the first theorem is set in the context of a Walrasian general equilibrium model, developed almost a hundred years after The *Wealth of Nations*. Since Smith wrote long before the modern mathematical language of economics was invented, he never rigorously stated, let alone proved, any version of the first theorem. That was first done by Lerner (1934), Lange (1942) and Arrow (1951).

To establish the first theorem, we need to sketch a general equilibrium model of an economy. Assume all individuals and firms in the economy are price takers: none is big enough, or motivated enough, to act like a monopolist. Assume each individual chooses his consumption bundle to maximize his utility subject to his budget constraint. Assume each firm chooses its production vector, or input–output vector, to maximize its profits subject to some production constraint. Note that we assume *self-interest*, or the absence of *externalities:* an individual cares only about his own utility, which depends only on his own consumption. A firm cares only about its own profits, which depend only on its own production vector.

The invisible hand of competition acts through prices; they contain the information about desire and scarcity that coordinate actions of self-interested agents. In the general equilibrium model, prices adjust to bring about equilibrium in the market for each and every good. That is, prices adjust until supply equals demand. When that has occurred, and all individuals and firms are maximizing utilities and profits, respectively, we have a competitive equilibrium.

The first theorem establishes that a competitive equilibrium is for the common good. But how is the common good defined? The traditional definition looks to a measure of total value of goods and services produced in the economy. In Smith, the 'annual revenue of the society' is maximized. In Pigou (1920), following Smith, the 'free play of self-interest' leads to the greatest 'national dividend'.

However, the modern interpretation of 'common good' typically involves Pareto optimality, rather than maximized gross national product. When ultimate consumers appear in the model, a situation is said to be Pareto optimal if there is no feasible alternative that makes everyone better off. Pareto optimality is thus a dominance concept based on comparisons of vectors of utilities. It rejects the notion that utilities of different individuals can be compared, or that utilities of different individuals can be summed up and two alternative situations compared by looking at summed utilities. When ultimate consumers do not appear in the model, as in the pure production framework to be described below, a situation is said to be Pareto optimal if there is no alternative that results in the production of more of some output, or the use of less of some input, all else equal. Obviously saying that a situation is Pareto optimal is not the same as saying it maximizes GNP, or that it is best in some unique sense. There are generally many Pareto optima. However, optimality is a common good concept that can get common assent: No one would argue that society should settle for a situation that is not optimal, because if A is not optimal, there exists a B that all prefer.

In spite of the multiplicity of optima in a general equilibrium model, most states are non-optimal. If the economy were a dart board and consumption and production decisions were made by throwing darts, the chance of hitting an optimum would be zero. Therefore, to say that

the market mechanism leads an economy to an optimal outcome is to say a lot. And now we can turn to a modern formulation of the first theorem:

First Fundamental Theorem of Welfare Economics Assume that all individuals and firms are self-interested price takers. Then a competitive equilibrium is Pareto optimal.

To illustrate the theorem, we focus on one simple version of it, set in a pure production economy. For a general versions of the theorem, with both production and exchange, the reader can refer to Mas-Colell et al. (1995).

In a general equilibrium production economy model, there are K firms and m goods, but, for simplicity, no consumers. We write $k = 1, 2, \ldots, K$ for the firms, and $j = 1, 2, \ldots, m$ for the goods. Given a list of market prices, each firm chooses a feasible input–output vector y_k so as to maximize its profits. We adopt the usual sign convention for a firm's input–output vector y_k : $y_{k\,j} < 0$ means firm k is a net user of good j, and $y_{kj} > 0$ means firm k is a net producer of good j. When we add the amounts of good j over all the firms, $y_{1j} + y_{2j} + \ldots + y_{Kj}$, we get the aggregate net amount of good j produced in the economy, if positive, and an aggregate net amount of good j used, if negative. What is feasible for firm k is defined by some fixed production possibility set Y_k. Under the sign convention on the input–output vector, if p is a vector of prices, firm k's profits are given by

$$\pi_k = p \cdot y_k.$$

A list of feasible input–output vectors $y = (y_1, y_2, \ldots, y_k)$ is called a *production plan* for the economy. A *competitive equilibrium* is a production plan $\hat{y}$ and a price vector p such that, for every k, $\hat{y}_k$ maximizes π_k subject to y_k's being feasible. (Since the production model abstracts from the ultimate consumers of outputs and providers of inputs, the supply equals demand requirement for an equilibrium is moot.)

If $y = (y_1, y_2, \ldots, y_k)$ and $z = (z_1, z_2, \ldots, z_K)$ are alternative production plans for the economy, z is said to *dominate* y if the following vector inequality holds:

$$\sum_k Z_k \geq \sum_k y_k.$$

The production plan y is said to be *Pareto optimal* if there is other production plan that dominates it. (Note that for two vectors a and b, $a \geq b$ means $a_j \geq b_j$ for every good j, with the strict inequality holding for at least one good.)

We now have the apparatus to state and prove the first theorem in the context of the pure production model:

First Fundamental Theorem of Welfare Economics, Production Version Assume that all prices are positive, and that $\hat{y}$, p is a competitive equilibrium. Then $\hat{y}$ is Pareto optimal.

To see why, suppose to the contrary that a competitive equilibrium production plan $\hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_k)$ is not optimal. Then there exists a production plan $z = (z_1, z_2, \ldots, z_K)$ that dominates it. Therefore

$$\sum_k Z_k \geq \sum_k \hat{y}_k.$$

Taking the dot product of both sides with the positive price vector p gives

$$p \cdot \sum_k Z_k \geq p \cdot \sum_k \hat{y}_k.$$

But this implies that, for at least one firm k,

$$p \cdot z_k > p \cdot \hat{y}_k$$

which contradicts the assumption that $\hat{y}_k$ maximizes firm k's profits. Q.E.D.

First Fundamental Theorem Drawbacks, and the Second Fundamental th

The first theorem of welfare economics is mathematically true but nevertheless open to objections. Here are the commonest.

First, the theorem is an abstraction that ignores the facts. Preferences of consumers are not given, they are created by advertising. The real economy is never in equilibrium, most markets are

characterized by excess supply or excess demand, and are in a constant state of flux. The economy is dynamic, tastes and technology are constantly changing, whereas the model assumes they are fixed. The cast of characters in the real economy is constantly changing, the model assumes it fixed.

Second, the theorem assumes competitive behaviour, whereas the real world is full of monopoly and market power.

Third, the theorem assumes there are no externalities. In fact, if in an exchange economy person l's utility depends on person 2's consumption as well as his own, the theorem does not hold. Similarly, if in a production economy firm *k's* production possibility set depends on the production vector of some other firm, the theorem breaks down. In a similar vein, the theorem assumes there are no public goods, that is, goods like national defence, judicial systems or lighthouses, that are necessarily non-exclusive in use. If such goods are privately provided (as they would be in a completely laissez-faire economy), then their level of production will be suboptimal.

Fourth, the theorem ignores distribution. Laissez-faire may produce a Pareto optimal outcome, but there are many different Pareto optima, and some are fairer than others. Some people are endowed with resources that make them extremely rich, while others, through no fault of their own, are extremely poor.

The first and second objections to the first theorem are beyond the scope of this article. The third, regarding externalities and public goods, is one that economists have always acknowledged. The standard remedies for these market failures involve various modifications of the market mechanism, including Pigouvian taxes (Pigou 1920) on harmful externalities, or appropriate Coasian legal entitlements to, for example, clean air (Coase 1960).

The important contribution of Pigou is set in a partial equilibrium framework, in which the costs and benefits of a negative externality can be measured in money terms. Suppose that a factory produces gadgets to sell at some market-determined price, and suppose that, as part of its production process, the factory emits smoke which damages another factory located downwind. In order to maximize its profits, the upwind factory will expand its output until its marginal cost equals price. But each additional gadget it produces causes harm to the downwind factory – the marginal external cost of its activity. If the factory manager ignores that marginal external cost, he will create a situation that is non-optimal in the sense that the aggregate net value of both firms' production decisions will not be as great as it could be. That is, what Pigou calls 'social net product' will not be maximized, although 'trade net product' for the polluting firm will be. Pigou's remedy was for the state to eliminate the divergence between trade and social net product by imposing appropriate taxes (or, in the case of beneficial externalities, bounties). The Pigouvian tax would be set equal to marginal external cost, and with it in place the gap between the polluting firm's view of cost and society's view would be closed. Optimality would be re-established.

Coase's contribution was to emphasize the reciprocal nature of externalities and to suggest remedies based on common law doctrines. In his view the polluter damages the pollutee only because of their proximity; for example, the smoking factory harms the other only if it happens to locate close downwind. Coase rejects the notion that the state must step in and tax the polluter. The common law of nuisance can be used instead. If the law provides a clear right for the upwind factory to emit smoke, the downwind factory can contract with the upwind factory to reduce its output, and if there are no impediments to bargaining, the two firms acting together will negotiate an optimal outcome. Alternatively, if the law establishes a clear right for the downwind factory to recover for smoke damages, it will collect external costs from the polluter, and thereby motivate the polluter to reduce its output to the optimal level. In short, a legal system that grants clear rights to the air to either the polluter or pollutee will set the stage for an optimal outcome, provided that bargaining is costless. If bargaining is costly, then the law should be designed with an eye towards minimizing social costs created by the externality.

With respect to public goods, since Samuelson (1954) derived formal optimality conditions for their provision, the issue has received much

attention from economists; one especially notable theoretical question has to do with discovering the strengths of people's preferences for a public good. If the government supplies a public judicial system, for instance, how much should it spend on it (and tax for it)? At least since Samuelson, it has been known that financing schemes like those proposed by Lindahl (1919), where an individual's tax is set equal to his marginal benefit, provide perverse incentives for people to misrepresent their preferences. Schemes that are immune to such misrepresentations (in certain circumstances) have been developed (Clarke 1971; Groves and Loeb 1975).

But it is the fourth objection to the first theorem that may be most fundamental. What about distribution?

There are two polar approaches to rectifying the distributional inequities of laissez-faire. The first is the command economy approach: a central bureaucracy makes detailed decisions about the consumption decisions of all individuals and production decisions of all producers. The main theoretical problem with the command approach is that it fails to create appropriate incentives for individuals and firms. On the empirical side, the experience of the late Soviet and Maoist command economies establish that highly centralized economic decision making leaves much to be desired, to put it mildly.

The second polar approach to solving distribution problems is to transfer income or purchasing power among individuals, and then to let the market work. The only kind of purchasing power transfer that does not cause incentive-related losses is the lump-sum money transfer. Enter at this point the standard remedy for distribution problems, as put forward by market-oriented economists, and our second major theorem.

The second fundamental theorem of welfare economics establishes that the market mechanism, modified by the addition of lump-sum transfers, can achieve virtually *any* desired optimal distribution. Under more stringent conditions than are necessary for the first theorem, including assumptions regarding quasi-concavity of utility functions and convexity of production possibility sets, the second theorem gives the following:

Second Fundamental Theorem of Welfare Economics Assume that all individuals and producers are self-interested price takers. Then almost any Pareto optimal equilibrium can be achieved via the competitive mechanism, provided appropriate lump-sum taxes and transfers are imposed on individuals and firms.

One version of the second theorem, restricted to a pure production economy, is particularly relevant to an old debate about the feasibility of socialism; see particularly Lange and Taylor (1939) and Lerner (1944). Anti-socialists including von Mises (1922) argued that informational problems would make it impossible tocoordinate production in a socialist economy; while pro-socialists, particularly Lange, argued that those problems could be overcome by a central planning board, which limited its role to merely announcing a price vector. This was called 'decentralized socialism'. Given the prices, managers of production units would act like their capitalist counterparts; in essence, they would maximize profits. By choosing the price vectors appropriately, the central planning board could achieve any optimal production plan it wished.

In terms of the production model given above, the production version of the second theorem is as follows:

Second Fundamental Theorem of Welfare Economics, Production Version Let $\hat{y}$ be any optimal production plan for the economy. Then there exists a price vector p such that $\hat{y}$, p is a competitive equilibrium. That is, for every k, $\hat{y}_k$ maximizes $\pi_k = p \cdot y_k$ subject to y_k being feasible.

The proof of the second theorem will not be presented here.

Adjusting the Economy and Voting

We rarely choose between a laissez-faire economy and a command economy. Our choices are almost always more modest. When choosing among alternative tax policies, or trade and tariff policies, or development policies, or anti-monopoly policies, or labour policies, or transfer policies, what shall guide the choice? The applied welfare

economist's advice is usually based on some notion of increasing total output in the economy. The practical political decision, in a democracy, is normally based on voting.

Applied Welfare Economics

The applied welfare economist usually focuses on ways to increase total output, 'the size of the pie', or at least to measure changes in the size of the pie. Unfortunately, theory suggests that the pie cannot be easily measured. This is so for a number of reasons. To start, any measure of total output is a scalar, that is, a single number. If the number is found by adding up utility levels for different individuals, illegitimate interpersonal utility comparisons are being made. If the number is found by adding up the values of aggregate net outputs of all goods, there is an index number problem. The value of a production plan will depend on the price vector at which it is evaluated. But, in a general equilibrium context, the price vector will depend on the aggregate net output vector, which will in turn depend on the distribution of ownership or wealth among individuals.

An early and crucial contribution to the analysis of whether or not the economic pie has increased in size was made by Kaldor (1939, p. 550), who argued that the repeal of the Corn Laws in England could be justified on the grounds that the winners might in theory compensate the losers: 'it is quite sufficient [for the economist] to show that even if all those who suffer as a result are fully compensated for their loss, the rest of the community will still be better off than before'. Unfortunately, Scitovsky (1941) quickly pointed out that Kaldor's compensation criterion (as well as one proposed around the same time by Hicks) was inconsistent. Consider a move from situation A to situation B. It is possible to judge B Kaldor superior to A (the move is an improvement) and simultaneously judge A Kaldor superior to B (the move back would also be an improvement). This Scitovsky paradox can be avoided via a two-edged compensation test, according to which B is judged better than A if (a) the potential gainers in the move from A to B could compensate the potential losers, and still remain better off, and (b) the potential losers could not bribe the gainers to forgo the move.

However, while Scitovsky's two-edged criterion has some logical appeal, it still has a major drawback: it ignores distribution. Therefore, it can make no judgement about alternative distributions of the same size pie. Even worse, both the Kaldor and the Scitovsky criteria would approve of a change that makes the wealthiest man in society richer by $1 billion, while making each of the million poorest people worse off by $999. This is a judgement that many people would reject as wrong or immoral.

Another important tool for measuring changes in economic welfare is the concept of consumer's surplus, which Marshall (1920, book 3, ch. 6) defined as the difference between what an individual would be willing to pay for an object, at most, and what he actually does pay. With a little faith, the economic analyst can measure aggregate consumers' surplus (note the new position of the apostrophe) by calculating an area under a demand curve, and this is in fact commonly done in order to evaluate changes in economic policy. The applied welfare economist attempts to judge whether the pie would grow in a move from A to B by examining the change in consumers' surplus (plus profits, if they enter the analysis). Some faith is required because consumers' surplus, like the Kaldor criterion, is theoretically inconsistent; see for example Boadway (1974).

Under certain circumstances, however, consumers' surplus inconsistencies can be ruled out. In particular, if individual utility functions are all quasilinear, of the form $u_i(x_i) = v_i(x_{ij}, j \neq m) + x_{im}$, then consumers' surplus paradoxes disappear. (Here $u_i(x_i)$ is person i's utility, as a function of his consumption bundle $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})$, and the utility function can be separated into two parts, the first one of which is a function $v_i(\cdot)$ which depends on quantities of all the goods except the mth, and the second of which is simply the quantity of the mth good. The mth good can be interpreted as the 'money' good; all the individuals like it, and value it the same way, with the same marginal utility, of one). The assumption of quasilinear preferences is a very

strong one if we think about 'real' commodities like wine and bread, but it has a certain intuitive appeal if we are inclined to believe in utility from 'money'.

To sum up this section, although the tools of applied welfare economics are widely used and very important in practice, in theory they should be viewed with some skepticism.

Voting

In many cases, interesting decisions about economic policies are made either by government bureaucracies that are controlled by legislative bodies, or by legislative bodies themselves, or by elected executives: in short, either directly or indirectly, by voting. The second theorem itself raises qsts about distribution that many would view as essentially political. How should society choose the Pareto-optimal allocation of goods that is to be reached via the modified competitive mechanism? How should the distribution of income be chosen? How can the best distribution of income be chosen from among many Pareto optimal ones? Majority rule is a commonly used method of choice in a democracy, both for political choices and economic ones, and we now turn our attention to it.

The practical objections to voting, the fraud, the deception, the accidents of weather, are well known. To quote Boss Tweed, the infamous 19th century chief of New York's Tammany Hall: 'As long as I count the votes, what are you going to do about it?' We will examine the theoretical problems.

The central theoretical problem with majority voting has been known since the time of Condorcet's *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, published in 1785: Voting may be logically inconsistent. The now standard Condorcet voting paradox assumes three individuals 1, 2 and 3, and three alternatives *x, y* and *z*, where the three voters have the following preferences:

$$\begin{array}{llll} 1: & x & y & z \\ 2: & y & z & x \\ 3: & z & y & x \end{array}$$

(Following an individual's number the alternatives are listed in his order of preference, from left to right). Majority voting between pairs of alternatives will reveal that *x* beats *y, y* beats *z,* and, paradoxically*, z* beats *x*.

It is now clear that such voting cycles are not peculiar; they are generic, particularly when the alternatives have a spatial aspect with two or more dimensions (Plott 1967; Kramer 1973). This can be illustrated by taking the alternatives to be different distributions of one economic pie. Suppose, in other words, that the distributional issues raised by the first and second theorems are to be 'solved' by majority voting, and assume for simplicity that what is to be divided is a fixed total of wealth, say 100 units.

Now let *x* be 50 units for person 1, 30 units for person 2 and 20 units for person 3. That is, let $x = (50, 20, 30)$. Similarly, let $y = (30, 50, 20)$ and $z = (20, 30, 50)$. The result is that our three individuals have precisely the voting paradox preferences. Nor is this result contrived, it turns out that *all* the distributions of 100 units of wealth are connected by endless voting cycles (see McKelvey 1976). The reader can easily confirm that for any distributions *u* and *v*, that he may choose, there exists a voting sequence from *u* to *v,* and another back from *v* to *u!*

The reality of voting cycles should give pause to the economist who recommends legislation about economic choices, especially choices among alternative distributions of income or wealth.

Social Welfare and the Third Fundamental th

How then can economic choices be made; how, for example, might the distribution problem be solved? One potential answer is to assert the existence of a Bergson (1938) economic welfare function $E(\cdot)$, that depends on the amounts of non-labour factors of production employed by each producing unit, the amounts of labour supplied by each individual, and the amounts of produced goods consumed by each individual. Then solve the problem by maximizing $E(\cdot)$, If necessary conditions for Pareto optimality are derived that

must hold for any $E(\cdot)$, this exercise is harmless enough; but if a particular $E(\cdot)$ is assumed and distributional implications are derived from it, then an objection can be raised: why that Bergson function $E(\cdot)$, and not a different one?

At first, in his modestly titled 'A Difficulty in the Concept of Social Welfare' (1950), and later, in his classic monograph *Social Choice and Individual Values* (1963), Kenneth Arrow brought together both the economic and political streams of thought sketched above. Arrow's theorem can be viewed in several ways: it is a statement about the distributional qsts raised by the first and second theorems; it is an extension of the Condorcet voting paradox; it is a statement about the logic of voting; and it is a statement about the logic of Bergson welfare functions, compensation tests, consumers' surplus tests, and indeed all the tools of the applied welfare economist. Because of its central importance, Arrow's theorem can be justifiably called the third fundamental theorem of welfare economics.

Arrow's analysis is at a high level of abstraction, and requires some additional model building. From this point onward we assume a given set of at least three distinct alternatives, which might be allocations in an exchange economy, distributions of wealth, tax bills in a legislature, or candidates in an election. The alternatives are written *x, y, z* and so on. We assume a fixed society of individuals, numbered 1, 2,. . ., n. Let *Ri* represent the preference relation of individual i, so xR_iy means person i likes x as well as or better than y. (Strict preference is shown with a P_i, and indifference with a I_i). A *preference profile* for society is a specification of preferences for each and every individual, or symbolically, $R = (R_1, R_2, \ldots, R_n)$. We write R_S for society's preference relation, arrived at in a way yet to be specified.

Arrow was concerned with the logic of how individual preferences are transformed into social preferences. That is, how is R_s found? Symbolically we can represent the transformation this way:

$$R_1, R_2, \ldots, R_n \rightarrow R_s.$$

Now if society is to make decisions regarding things like distributions of wealth, it must 'know' when one alternative is as good as or better than another, even if both are Pareto optimal. To ensure it can make such decisions, Arrow requires that R_S be *complete*. That is, for any alternatives x and y, either xR_Sy or yR_Sx (or both, if society is indifferent between the two). If society is to avoid the illogic of voting cycles, its preferences ought to be *transitive*. That is, for any alternatives *x, y* and *z,* if xR_Sy and yR_Sz, then xR_Sz. Following Sen (1970), we call a transformation of preference profiles into complete and transitive social preference relations an *Arrow social welfare function*, or more briefly, an Arrow SWF.

Anyone can make up an Arrow SWF, just as anyone can make up a Bergson function, or for that matter a simple moral judgement about when one distribution of wealth is better than another. But arbitrary judgments are unsatisfactory and so are arbitrary Arrow functions. Therefore, Arrow imposed some reasonable conditions on his function. Following Sen's (1970) version of Arrow's theorem, there are four conditions:

1. Universality. The function should always work, no matter what individual preferences might be. It would not be satisfactory, for example, to require unanimous agreement among all the individuals before determining social preferences.
2. Pareto consistency. Social preferences should be consistent with the *Pareto criterion*. That is, if everyone prefers x to y, then the social preference is x over y.
3. Independence. Suppose there are two alternative preference profiles for individuals in society, but suppose individual preferences regarding x and y are exactly the same under the two alternatives. Then the social preference regarding x and y must be exactly the same under the two alternatives. In particular, if individuals change their minds about a third 'irrelevant' alternative, this should not affect the social preference regarding x and y.
4. Non-dictatorship. There should not be a dictator. In Arrow's abstract model, person i is a *dictator* if society always prefers what he prefers; that is, if i prefers x to y, then the social preference is x over y.

An economist or policymaker who wants an ultimate answer to qsts involving distributions, or qsts involving choices among alternatives that are not comparable under the Pareto criterion, could use an Arrow SWF for guidance. Unfortunately, Arrow showed that imposing conditions 1 to 4 guarantees that Arrow functions *do not exist:*

Third Fundamental Theorem of Welfare Economics There is no Arrow social welfare function that satisfies the conditions of universality, Pareto consistency, independence, non-dictatorship.

In order to illustrate the logic of the theorem, we will use a stronger assumption than independence. This assumption is called *neutrality–independence–monotonicity* (NIM), defined as follows. Suppose for some group of individuals *V*, some preference profile, and some pair of alternatives *x* and *y*, all members of *V* prefer *x* to *y*, all individuals not in *V* prefer *y* to *x*, and the social preference is *x* over *y*. Then for any preference profile and any pair of alternatives *w* and *z,* if all people in *V* prefer *w* to *z,* the social preference must be *w* over *z*. In short, if *V* gets its way in one instance, when everyone opposes it, then it must have the power to do it again, when the opposition may be weaker.

A group of individuals *V* is said to be decisive if for all alternatives *x* and *y*, whenever all the people in *V* prefer *x* to *y*, society prefers *x* to *y.* Note that if person *i* is a dictator, {i} is decisive, and conversely. Assumption NIM asserts that, if *V* prevails when it is opposed by everyone else, it must be decisive. If the social choice procedure is majority rule, for example, any group of $(n + 1)/2$ members, for n odd, or $(n/2) + 1$ members, for n even, is decisive. Moreover, it is clear that majority rule satisfies the NIM assumption, since if *V* prevails for a particular *x* and *y* when everyone outside of *V* prefers *y* to *x*, then *V* must be a majority, and must always prevail. (Majority rule is just one example of a procedure that satisfies NIM; there are many other procedures that also do so.)

Now we are ready to turn to a short and simple version of the third theorem.

Third Fundamental Theorem of Welfare Economics, Short Version There is no Arrow SWF that satisfies the conditions of universality, Pareto consistency, neutrality–independence–monotonicity, and non-dictatorship.

The proof goes as follows. First, there must exist decisive groups of individuals, since by the Pareto consistency requirement the set of all individuals is one. Now let *V* be a decisive group of minimal size. If there is just one person in *V*, he is a dictator. Suppose then that *V* includes more than one person. We show this leads to a contradiction.

If there are two or more people in *V*, we can divide it into non-empty subsets *V*1 and *V*2. Let *V*3 represent all the people who are in neither *V*1 nor *V*2. (*V*3 may be empty). By universality, the Arrow function must be applicable to any profile of individual preferences. Take three alternatives *x*, *y* and *z* and consider the following preferences regarding them:

$$\begin{array}{llll} \text{For individuals in } V_1: & x & y & z \\ \text{For individuals in } V_2: & y & z & x \\ \text{For individuals in } V_3: & z & x & y \end{array}$$

(At this point the close tie between Arrow and Condorcet is clear, for these are exactly the voting paradox preferences!)

Since *V* is by assumption decisive, *y* must be socially preferred to *z,* which we write yP_Sz. By the assumption of completeness for the social preference relation, either xR_Sy or yP_Sx must hold. If xR_Sy holds, since xR_Sy and yP_Sz, then xP_Sz must hold by transitivity. But now *V*1 is decisive by the NIM assumption, contradicting *V's* minimality. Alternatively, if *yPSx* holds, *V*2 is decisive by the NIM assumption, again contradicting *V's* minimality. In either case, the assumption that *V* has two or more people leads to a contradiction. Therefore *V* must contain just one person, who is, of course, a dictator! Q.E.D.

Since the third theorem was discovered, a whole literature of modifications and variations has been spawned. But the depressing conclusion has remained more or less the same: there is no logically infallible way to aggregate the preferences of diverse individuals into a social preference relation. Therefore, there are no logically infallible ways to vote, or to solve the problems of distribution of income and wealth in society.

Social Welfare After Arrow

Social Choice Functions and Strategy

The third fundamental theorem of welfare economics tells us that we cannot find an Arrow social welfare function satisfying certain reasonable requirements. An Arrow function maps preference profiles (that is, preference relations for each and every member of society) into social preference relations. But in order to make judgements about what alternative is *best* for society, it is not really necessary to have a social preference relation. Suppose we just had a rule that tells us, if the set of alternatives is *x, y, z* and so on, and the preference profile is $R = (R_1, R_2, \ldots, R_n)$then the best alternative is such-and-such? Such a rule would be a mapping from preference profiles into alternatives, written symbolically as follows:

$$R_1, R_2, \ldots, R_n \rightarrow x.$$

Such a rule is called a *social choice function*, or SCF for short. An Arrow function produces a social ranking of all the alternatives; an SCF, in contrast, just produces a winner. As an example, think of plurality voting, with some kind of rule to break ties.

The essential difficulty with SCFs is that they may create obvious incentives for people to misrepresent their preferences, so as to obtain better (for them) social choices. As an example, consider again the Condorcet voting paradox preferences:

$$\begin{array}{llll} 1: & x & y & z \\ 2: & y & z & x \\ 3: & z & x & y \end{array}$$

Suppose the three people use plurality voting (each person casts one vote for his favorite), and, in case of a tie, the social choice is the outcome closest to the beginning of the alphabet. Under this rule, if 1 votes for his favourite, *x,* and persons 2 and 3 do likewise, there is a three-way tie, which is resolved with the (alphabetical) choice of *x*. Now put yourself in the shoes of person 2. You will immediately see that, if persons 1 and 3 continue to vote for their favourites, and if you switch from your favourite *y* to your second favourite *z*, then social choice changes, from *x* to *z*, making you better off! You are in effect being asked 'what is your preference relation?' Instead of answering honestly (*y z x*), you offer, in effect, a false preference relation (*z y x*).

Reporting a false preference relation in order to bring about an SCF outcome that you prefer to the one you get if you are honest, is called *strategic behaviour*, or *strategizing*. It is obviously a bad thing if an SCF produces lots of opportunities for strategic behaviour: if individuals are commonly strategizing, there is no reason to believe that the outcome, based as it is on false reports, is truly best for society. If an SCF has the property that it is never advantageous for anyone to report a false preference relation it is called *strategy-proof*. For instance, suppose an SCF always chooses the alternative that is first in the alphabetical list of alternatives. This SCF would be frustrating and idiotic, but it would be strategy-proof.

The Gibbard–Satterthwaite th

This leads to a natural qst: are there SCFs that are immune to strategic behaviour, and that satisfy a few other reasonable conditions? Note that the question is very similar in style to the question that Arrow asked about Arrow SWFs. What would the reasonable conditions be? First (similar to Arrow), the SCF ought to be *universal*; that is, it should work no matter what the profile of individual preferences might be. Second (also similar to Arrow), there should be no *dictator*. In the SCF context person *i* is a dictator if the social choice is always a top-ranked alternative for person *i*. Third (and different from Arrow), the SCF should be *non-degenerate*. This means that, for any alternative *x,* there must be some preference profile which would give rise to *x's* being the social choice. (This requirement excludes the SCF that always chooses the first alternative in the alphabetical list). Now we can ask the qst: do there exist SCFs which are universal, non-degenerate, strategy-proof and non-dictatorial?

This question was asked and answered, independently, by Gibbard (1973) and Satterthwaite (1975). The Gibbard–Satterthwaite result turns out to be logically very close to third fundamental theorem of welfare economics; in fact, Gibbard

uses Arrow's theorem to prove his theorem, and Satterthwaite shows that his theorem can be used to prove Arrow's. Following is the Gibbard–Satterthwaite theorem. The proof is omitted; a simplified and restricted version of the theorem, and a simple proof, can be found in Feldman and Serrano (2006):

Gibbard–Satterthwaite th There is no social choice function that satisfies the conditions of universality, non-degeneracy, strategy-proofness and non-dictatorship.

Like the third fundamental theorem, the Gibbard–Satterthwaite theorem is starkly negative; it says that, if you want a decision-making process, an SCF to be precise, that has desirable characteristics, including being immune to strategic manipulation, you are bound to be disappointed. To put it differently, for any reasonable SCF, there will be circumstances under which some person will want to falsely report his preferences, resulting in a perversion of the process, and an outcome that may not be desirable for society.

If a decision-making process works in a way that offers each individual *no* incentive to misrepresent his preferences, no matter what preferences the other $n-1$ individuals might be reporting, we say that honestly reporting one's preferences (or telling the truth) is a *dominant strategy*. The Gibbard–Satterthwaite result then says that, if a social choice function satisfies the conditions of universality, non-degeneracy and non-dictatorship, truth-telling will not be a dominant strategy. That is, there will be some reported preference relations of all individuals except *i,* which will provide an incentive for individual *i* to lie. If everyone else is saying such-and-such (which might be true or false), person *i* will give a false report. This is what strategy-proofness excludes.

But what if we narrowed this broad notion of strategy-proofness; what if we required that *i* not have an incentive to lie when the others are reporting the truth, rather than requiring that *i* never have an incentive to lie, no matter what the others are reporting?

Implementation and the Maskin th

If telling the truth is a best strategy when others are telling the truth, rather than always, then truth telling is a *Nash strategy*, rather than a *dominant strategy*. The theory of implementation, or mechanism design, provides a way out the negativity of the Gibbard Satterthwaite theorem; it provides a way to implement an SCF, or support its choices, by incorporating truth telling about preferences into Nash equilibrium strategies in games.

The major theorem on implementation is due to Maskin (1999), whose paper first circulated in 1977. In Maskin's model, there is a social planner (or central authority) who wants to bring about choices according to a given SCF, which we now call *F*. The planner knows *F*, as do all the members of society. Given any preference profile *R*, the SCF produces an outcome $F(R) = x$. But the planner does not know the true preferences of the individuals. He must rely on the individuals to report their preferences, and they may lie. We assume for simplicity that every individual knows the true preference relation for himself *and* every other individual; that is, each *i* knows the true preference profile, but the social planner doesn't. (This is obviously a strong assumption). From this point on, when one preference profile may be true and another may be false, we will use the unadorned *R* to represent the *true* profile. The social planner receives reports on preferences, or preference profiles, from the individuals, but they may be lies. We let $\hat{R}_i$ represent a reported preference relation for person *i,* which may be false; similarly $\hat{R} = \left(\hat{R}^i{}_1, \hat{R}_2, \ldots, \hat{R}_n\right)$ represents a reported preference *profile*, which may be false; and $\hat{R}^i = \left(\hat{R}^i{}_1, \hat{R}^i{}_2, \ldots, \hat{R}^i{}_n\right)$ represents a preference *profile, reported by person i*, which may false. The social planner wants to devise a method, a mechanism, to induce individuals to honestly report preferences. That way he will get hold of the true preference profile *R*, and produce the desired outcome $F(R) = x$.

How might this be done? The intuition is to ask each and every individual to report a preference profile. (Note that, since we assume all the individuals know each other's true preferences, it is no more challenging for an individual to report a preference profile, comprising preference relations for everyone in society, than it is to report his own preference relation). If all the reported preferences profiles agree, there's a good chance

they are all true, and the planner might accept the generally agreed-upon profile. If they all agree except for one, the one that's out of line probably comes from a liar, and he should be given a motive to avoid lying. (If the social planner were a despot, the out-of-line person would be shot. Note also that there must be three or more individuals in society to discover whose report is out of line). Finally, when the reported preference profiles generally disagree, the social planner needs a way to avoid having the process stop at an inappropriate Nash equilibrium.

Let us be more precise. Maskin's algorithm for implementing an SCF works as follows. Each person i reports a message m_i, which is composed of an alternative x, a preference profile $\hat{R}^i$, and a non-negative integer. (i) If every message is the same, of the form $\left(x = F\left(\hat{R}\right)\hat{R}, 0\right)$ then the social planner chooses x. (ii) If every message but one is the same, of the form $\left(x = F\left(\hat{R}\right)\hat{R}, 0\right)$; for every person but j, while j reports a message of the form $(\mathrm{y} \neq \mathrm{x}, \hat{R}^j,$ anything), then the social planner chooses y, unless person j would like x *less than* y according to $\hat{R}_j$, the person-j preference relation that all the other people are reporting. If this is the case the planner chooses x. (Person j is not shot. He simply does not gain, and may lose, from his deviation). (iii) In all other cases, the social planner finds the person who proposes the highest integer (with some method for resolving ties), and chooses the alternative named by that person.

Now the qsts can be framed. First, given this mechanism, would $m_1 = m_2 = \ldots = m_n = (F(R), R, 0)$, with R the true preference profile, constitute a Nash equilibrium? Second, if $(m_1, \ldots, m_n)$ is any Nash equilibrium list of messages in this mechanism, can we be sure the resulting chosen alternative will be $F(R)$?

The answers are 'yes' and 'yes', under certain general assumptions. The assumptions of Maskin's theorem are as follows. First, there must be three or more individuals (so that a deviant message can be spotted). Second, a mild diversity condition must be satisfied. Maskin uses a condition called *no veto*. Loosely speaking, this means that, if $n - 1$ people prefer x to all the other alternatives, then the SCF must choose x. Alternatively, one can assume the existence of a private economic good, that everyone values. This guarantees that individuals will disagree about what alternatives are best. In this article we will simply assume *diversity*, meaning the following: for any given alternative x, there exist at least two people, each of whom prefers something else to x.

Third, the social choice function F must satisfy an intuitive condition called *Maskin monotonicity.* (The condition is actually a distant relative of the NIM assumption used in the simple version of Arrow's theorem presented above). Maskin monotonicity means the following. Let $\hat{R}$ and R be any two preference profiles. (These may be true or false; it does not matter in this context). Suppose $F\left(\hat{R}\right) = x$, and suppose that, for all individuals i and all alternatives y, $x\hat{R}_i y$ implies $xR_i y$. Then $F(R)=x$. (In other words, in a hypothetical transition from $\hat{R}_i$ to Ri, for every person i the set of alternatives that i likes less than x or the same as x has expanded, or at least hasn't shrunk. Since x was the social choice under $\hat{R}$, x must continue to be the social choice under R). With these three conditions, Maskin proved:

Maskin Th Assume $n \geq 3$. Assume diversity and Maskin monotonicity. Then the mechanism described above implements the SCF F, in the sense that truthful messages leading to $F(R)$ comprise a Nash equilibrium, and in the sense that any Nash equilibrium list of messages results in the social planner choosing $F(R)$.

We will not provide all of the proof, but the logic is as follows. First, establish that $m_1 = m_2 = \ldots = mn = (F(R), R, 0)$is a Nash equilibrium, where R is the true preference profile. This is rather obvious, given rules (i) and (ii) of the Maskin algorithm. Second, establish that under rules (ii) and (iii), there are no Nash equilibria. This follows rather easily from the diversity assumption. Third, establish that, if $m_1 = m_2 = \ldots = mn = \left(F\left(\hat{R}\right)\hat{R}, 0\right)$is any Nash equilibrium, then $F\left(\hat{R}\right) = F(R)$. That is, given a Nash equilibrium based on a universally reported, but possibly false, preference profile, the outcome implemented is the same as if the true preference profile had been reported. This follows from the assumption of Maskin monotonicity.

Maskin also provided a near converse this theorem, which says that Maskin monotonicity is a necessary condition for any SCF F to be implementable. Relatively simple proofs of both Maskin theorems are available in Feldman and Serrano (2006).

Last Words

Where does welfare economics now stand? The first and second theorems are encouraging results that suggest the market mechanism has great virtue: competitive equilibrium and Pareto optimality are firmly bound. The third theorem exposes impossibilities and paradoxes in economic choices, voting choices, and, in general, almost any choices made collectively by society. The Gibbard–Satterthwaite theorem, like the third theorem, is a starkly negative result: any plausible social choice function will, under some circumstances, produce incentives for someone to lie. But the Maskin theorem is a ray of hope; it suggests a way for a social planner to design a game, whose Nash equilibria will implement a desired social choice function.

See Also

- Arrow's Theorem
- Compensation Principle
- Consumer Surplus
- Externalities
- General Equilibrium
- Mechanism Design
- Social Choice

Bibliography

Arrow, K.J. 1950. A difficulty in the concept of social welfare. *Journal of Political Economy* 58: 328–346.

Arrow, K.J. 1951. An extension of the basic theorems of classical welfare economics. In *Second Berkeley symposium on mathematical statistics and probability*, ed. J. Neyman. Berkeley: University of California Press.

Arrow, K.J. 1963. *Social choice and individual values*. 2nd ed. New York: John Wiley and Sons.

Bergson, A. 1938. A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52: 310–334.

Boadway, R. 1974. The welfare foundations of cost–benefit analysis. *Economic Journal* 84: 926–939.

Clarke, E.H. 1971. Multipart pricing of public goods. *Public Choice* 11: 17–33.

Coase, R.H. 1960. The problem of social cost. *Journal of Law and Economics* 3: 1–44.

Feldman, A.M., and R. Serrano. 2006. *Welfare economics and social choice theory*. 2nd ed. New York: Springer.

Gibbard, A. 1973. Manipulation of voting schemes: A general result. *Econometrica* 41: 587–601.

Groves, T., and M. Loeb. 1975. Incentives and public inputs. *Journal of Public Economics* 4: 211–226.

Kaldor, N. 1939. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal* 49: 549–552.

Kramer, G.H. 1973. On a class of equilibrium conditions for majority rule. *Econometrica* 41: 285–297.

Lange, O. 1942. The foundations of welfare economics. *Econometrica* 10: 215–228.

Lange, O., and F.M. Taylor. 1939. *On the economic theory of socialism*. Minneapolis: University of Minnesota Press.

Lerner, A.P. 1934. The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies* 1: 157–175.

Lerner, A.P. 1944. *The economics of control*. New York: The Macmillan Company.

Lindahl, E. 1919. Just taxation – A positive solution. Trans. and reprinted In *Classics in the theory of public finance*, ed. R.A. Musgrave and A.T. Peacock. New York: Macmillan, 1958.

Marshall, A. 1920. *Principles of economics*. 8th ed. London: Macmillan.

Mas-Colell, A., M.D. Whinston, and J.R. Green. 1995. *Microeconomic theory*. New York: Oxford University Press.

Maskin, E. 1999. Nash equilibrium and welfare optimality. *Review of Economic Studies* 66: 23–38.

McKelvey, R. 1976. Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory* 12: 472–482.

Pigou, A.C. 1920. *The economics of welfare*. London: Macmillan, Part II.

Plott, C.R. 1967. A notion of equilibrium and its possibility under majority rule. *American Economic Review* 57: 787–806.

Samuelson, P.A. 1954. The pure theory of public expenditure. *Review of Economics and Statistics* 36: 387–389.

Satterthwaite, M.A. 1975. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10: 187–217.

Scitovsky, T. 1941. A note on welfare propositions in economics. *Review of Economic Studies* 9: 77–88.

Sen, A.K. 1970. *Collective choice and social welfare*. San Francisco: Holden-Day.

von Mises, L. 1922. *Socialism: An economic and social analysis*. 3rd ed., and Trans. Indianapolis: Liberty Classics, 1981.

Welfare State

Assar Lindbeck

Abstract

This article starts out with a brief discussion of the historical background, the justifications and the political forces behind the creation of the modern welfare state. It also summarizes its major achievements in terms of economic efficiency and redistribution. The article also tries to identify some major problems of contemporary welfare-state arrangements, distinguishing problems caused by exogenous shocks from those related to endogenous behaviour adjustments by individuals to the welfare state itself. The latter include tax distortions, moral hazard, and endogenous changes in social norms concerning work and benefit dependency.

Keywords

Advantageous selection; Adverse selection; Aging and retirement; Altruism; Baumol, W.; Birth rates; Child care; Civil society; Dependency; Direct provision vs subsidy; Distribution of income; Early retirement pensions; Education; Employment subsidies; Enlightened self-interest; Family; Free riding; Gini coefficient; Globalization; Health care; Human capital investment; Human services; Immigration; Income effects; Income insurance; Income smoothing; Individualization; Industrialization; Intergenerational contracting; Intergenerational transfers; International trade (theory); Intra-individual redistribution; Job-security legislation; Labour supply; Labour-Intensive human services; Labour-Market programmes; Leisure; Luxembourg Income Study; Marginal tax rates; Market failure; Moral hazard; Myopia; Non-governmental organizations; Non-profit organizations; Old-age care; Political economy of income redistribution; Preschooling; Productivity growth; Public social spending; Redistribution of income and wealth; Retraining; Risk sharing; Sickness benefits; Social insurance; Social norms; Structural change; Structural unemployment; Substitution effect; Sweden; Tax base; Tax competition; Tax distortions; Tax wedge; Transfer payments; Unemployment benefits; Universal vs. targeted transfers; Urbanization; User fees; Voucher systems; Welfare state; Working poor

JEL Classifications

H1

According to a narrow definition, the welfare state comprises two types of government spending arrangements: (a) cash benefits to households (transfers, including mandatory income insurance) and (b) subsidies or direct government provision of human services (such as child care, preschooling, education, health care and old-age care). By broader definitions, the welfare state may also include price regulation (such as rent control and agricultural price support), housing policies, regulation of the work environment, job-security legislation, and environmental policies. This article is confined to the narrow definition.

Across developed Organisation for Economic Co-operation and Development (OECD) countries, total welfare-state spending ('public social spending'), including spending on education, varies today (2006) from about a fifth of GDP to about a third. As we would expect, the share is tightly related to the degree of 'universality' of public social spending, that is, the extent to which benefits are extended to individuals in all income classes rather than largely targeted to particular groups of individuals, such as low-income groups. Broadly speaking, the lowest figures are currently found in Anglo-Saxon countries, while the highest appear in the Nordic countries – with other countries in Western Europe somewhere in-between. Indeed, nowadays welfare states are usually classified in the context of such geographical clusters rather than according to distinctions between Bismarck- and Beveridge-type welfare states, or distinctions in terms of ideological categories along the lines suggested by Esping-Andersen (1990).

Justifications and Explanations

Urbanization has diminished the reliability of the family as a basis for reallocating income (or consumption) over the individual's life cycle, reducing income risk, and providing human services. Moreover, in connection with industrialization, new types of labour contracts emerged according to which unemployment and retirement became more discrete (abrupt) events than earlier (Atkinson 1991). Industrialization and the subsequent increase in office work also required an expansion of education at all levels. Meanwhile, progress in health and medicine enhanced the usefulness of professional medical services.

Needless to say, such developments by themselves do not justify government intervention in the fields of income insurance and human services, rather than simply leaving people to rely on voluntary solutions via markets and private networks ('civil society'). There are, however, well-known efficiency and distributional justifications for government intervention in these fields (see, for instance, Barr 1998). It is useful to divide the *efficiency justifications* into three categories.

First, the microeconomic literature identifies a number of limitations ('failures') in markets for voluntary income insurance: advantageous selection ('cream skimming') of insurance applicants, when insurance providers can differentiate between low-risk and high-risk individuals; adverse selection, when insurance providers cannot do so; myopia, when individuals underestimate their future income needs; and free riding on the altruism of others, when individuals expect others to assist them in case of economic distress. Mandatory income insurance ('social insurance') helps solve all these problems. Moreover, poor individuals may simply believe that they cannot afford to save or to buy income insurance: their marginal evaluation of immediate consumption is higher than their marginal evaluation of future income security. Paternalistic governments may prefer to deal with this issue by mandatory insurance rather than by cash transfers to such individuals. In addition, a monopoly provider may largely avoid marketing costs – although at the expense of individuals' freedom of choice.

Even if some of these problems may also be mitigated by group insurance, such arrangements are associated with well-known weaknesses. For instance, occupational income insurance often results in limited portability across jobs, and sometimes deficient financial viability, in particular when individual production firms or industries are in charge of the programmes. In some countries, however, such problems are avoided by institutional integration of occupational and governmentoperated arrangements ('corporatist' systems), such as in Germany and France.

Second, mandatory income insurance may also bring about risk sharing across generations. This is difficult to achieve by voluntary contracts alone since the potential parties of such contracts may not live simultaneously – both when the contract is signed and when it is supposed to be fulfilled.

Third, economists generally agree that investment in human capital (such as education and health care) tends to be suboptimal without government intervention (in the form of subsidies or direct government provision), either because of the difficulties in borrowing with expected future human capital as collateral or because of unexploited (positive) externalities in connection with such investment.

While the efficiency gains from government intervention in the context of the first two justifications show up in improved income smoothing and risk sharing, the efficiency gain according to the third justification takes the form of higher labour productivity and/or faster economic growth – provided disincentives due to higher government spending do not dominate these potential efficiency gains.

The *distributional justifications* for welfare-state arrangements also appear in different forms.

First, in the case of policies designed to fight poverty, it is natural to refer to genuine altruism or enlightened self-interest (a desire to mitigate negative externalities, such as ugly neighbourhoods and street crime). Intergenerational transfers in favour of old cohorts – for instance, via a pay-as-you-go (paygo) pension system – may also be justified by altruism, since lifetime income tends to be lower for older cohorts than for subsequent cohorts in growing economies.

Second, income insurance automatically reduces the overall dispersion of the *ex post* distribution of income. This holds for both yearly and lifetime income. Moreover, social insurance, as usually designed, may often reduce the *ex ante* dispersion of the distribution since such arrangements are seldom actuarially fair. A fairly common belief is that increased income security, and perhaps also a reduction in the overall dispersion of the distribution of income, up to a point tends to promote social peace, and that this in turn is favourable for economic growth; indeed, there is some empirical evidence in support of this view (Alesina and Rodrik 1994). In other words, a distributional argument may, up to a point, be turned into an efficiency justification for income insurance and redistribution of lifetime income.

Of course, neither historical background factors nor theoretical justifications (rationales) by themselves can *explain* the actual emergence and expansion of welfare-state arrangements. References to the political processes are required. In countries where policies are based on electoral processes, the distribution of voting power across socio-economic groups is a natural starting point. It is also tempting to explain politically generated redistribution across generations by the distribution of voting power across cohorts. For instance, current generations may transfer resources to themselves at the expense of future generations, which (by definition) do not have voting rights, although they may later renege on political favours acquired by earlier generations. At the same time, young adults with children would be expected to push politically for education (and infrastructure investment), while older cohorts are particularly likely to push for paygo pension systems and old-age care. The political outcome of such diverse interests, then, would be expected to depend on the relative power of different cohorts.

Indeed, some authors have tried to explain the emergence of modern social spending in western countries from the mid-19th century to the early 20th century by the gradual widening of franchise (Flora and Alber 1981; Lindert 2004). This is no doubt a realistic hypothesis. There are, however, obvious limitations to policy explanations in terms of relative voting powers of different interest groups. In the late 19th century (and even earlier), some welfare-state arrangements actually emerged in favour of individuals without voting rights; important examples include poor relief, mandatory and subsidized (even free) primary education, work-injury insurance and modest pensions. It is, therefore, tempting to assume that altruism and enlightened self-interest also help explain early welfare-state reforms – another example of how a justification may be turned into an explanation of actual development.

Moreover, the main expansion of welfare-state spending did not take place until half a century after the emergence of general franchise – indeed, mainly during the first three decades after the Second World War. One explanation for this apparent time lag may be that urbanization and industrialization were gradual processes, so that the political demand for new social arrangements likewise emerged gradually. It may also have taken considerable time to mobilize new groups of eligible voters. The time lags, and related gradualism in the expansion of welfare-state spending, could perhaps also be regarded as the result of an 'experimental approach' on the part of politicians or voters, due to uncertainty about the effects in various dimensions of higher welfare-state spending and related tax increases.

Achievements

Not only the level but also the composition of welfare-state spending, such as between transfers and human services, differs across countries. For instance, while about half of total public social spending consists of transfers in Western Europe as a whole (varying from 33 per cent in Iceland to 60 per cent in Austria), the corresponding figure is about 42 per cent in Anglo-Saxon countries outside Europe.

Transfers

What, then, is the relation between the size of aggregate government transfers on the one hand, and the degree of income security and government-induced redistribution of income

Welfare State, Table 1 Composition of total public social expenditures in 2001 (% GDP)

	United States			Western Europe[a]		
	Public	Private	Total	Public	Private	Total
Cash transfers	7.9	4.3	12.2	14.2	1.8	16.0
Pensions	*6.1*	*3.8*	*9.9*	*8.5*	*1.0*	*9.5*
Human services	11.9	7.2	19.1	15.1	0.9	16.0
Health	*6.2*	*5.0*	*11.1*	*6.4*	*0.4*	*6.8*
Education	*5.1*	*2.3*	*7.3*	*5.4*	*0.4*	*5.8*
Active labour market programmes	*0.1*		*0.1*	*0.9*		*0.9*
Total social expenditure	19.8	11.6	31.3	29.3	2.7	32.0

Notes: [a]Unweighted averages have been calculated for Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom. Figures for private health spending only cover private insurance programmes and exclude individual private health costs
Sources: Adema and Ladaique (2005), OECD (2004)

across households on the other? To answer the first aspect of the question, it is important to consider the extent to which government-provided arrangements are a substitute for private income insurance. To answer the second aspect of the question, we would ideally also need to determine the extent to which government transfers have resulted in induced (endogenous) changes in the distribution of factor income (general equilibrium effects). Unfortunately, our knowledge on both issues is quite limited.

Scattered evidence suggests, however, that voluntary private income insurance and social insurance are rather close substitutes at the margin. In particular, government-provided benefits tend to be topped up by occupational pensions in countries with only modest public benefits (Pearson and Martin 2005, pp. 8–10). As a result, *total* yearly per capita disposable income of retirees does not differ much across the eight west European countries studied by Forssell et al. (2000), in spite of considerable differences in the replacement rates in government-operated pension systems. It is also noticeable that total (public plus private) pensions are at least as large as a share of GDP in the United States as in western Europe (indeed, they are somewhat larger in the United States) in spite of the fact that the GDP share of public pensions is higher in western Europe, and that the population is younger in the United States (Table 1). Another example is that *total* per capita sick-pay benefits do not vary much among six west European countries studied by Kangas and Palme (1993), in spite of quite different replacement rates in government-operated systems – although the substitution is not complete.

There seems to be less substitution between public and private provision in countries where there is no government-operated system at all. The relatively low coverage of sick-pay insurance, sick-care insurance, and paid parental leave ('parent insurance') in the United States is a suggestive illustration. Thus, in areas where there is no government-operated system at all, it seems that the earlier discussed obstacles to the emergence of voluntary insurance arrangements 'kick in'.

Since the distribution of disposable income is considerably more even than the distribution of factor income, it is natural to argue that welfare-state arrangements, and their financing, actually contribute to reducing the unevenness of the distribution of income. Moreover, based on data from the Luxembourg Income Study (LIS), Korpi and Palme (1998) found that the relative difference between the market-income Gini coefficient and the disposable-income Gini coefficient tends to be larger in countries with universal transfer systems than in those with a strongly targeted system. (Market income is then defined as factor income plus occupational pensions.) In this sense, universal systems tend in fact to be more redistributive than targeted systems. However, this conclusion does not hold concerning the redistribution *per unit* of aggregate public social spending; rather, the reverse tends to be the case (although the difference is not statistically significant).

The observation that welfare-state arrangements, in fact, have reduced the dispersion of the distribution of yearly income relies, of course, on the implicit assumption that an induced widening of the distribution of factor income has not offset the direct impact on the distribution of disposable income. One indicator that such adjustments have not taken place is that the distribution of yearly factor income did not become more uneven – at least not much – during the period when the generosity of public benefit systems increased the most, that is, from the late 1940s to the mid-1970s. Moreover, the subsequent widening of the distribution of yearly factor income in a number of countries (until about the mid-1990s) has been particularly pronounced in the United States and the United Kingdom, that is, in countries where welfare-state spending has increased *less* than in other countries. Thus, it seems reasonable to assume that government transfer systems (including social insurance) have, in fact, reduced the dispersion of the distribution of yearly disposable income.

Human Services

In most developed countries, government intervention in the area of human services mainly takes the form of direct provision rather than general subsidies of such services. The effects of these policies would, however, be expected to differ systematically between low- and high-income citizens. One reason is that the per capita volume (or quality) provided by the government is often larger than what low-income individuals would have chosen themselves. Since human services cannot be resold in the market, the consumption of such services would be expected to increase among low-income groups. By contrast, it would be expected to fall among high-income groups, on the realistic assumption that human services, in contrast to income-insurance cash benefits, are difficult to supplement. (For instance, as a rule, parents do not divide up their children's attendance between a public and a private childcare centre or school.) Such a fall in consumption of human services among high-income groups would also be expected to take place among individuals who abstain from the public services offered and instead buy their services in the market. The reason, of course, is that their disposable income is reduced by the taxes they have to pay to finance the provision of human services to other citizens (basically reflecting an income effect).

There is a corollary to this reasoning: unless the volume provided is quite large, it is probably easier for the government to control the distribution than the aggregate volume of human services by direct provision. Total per capita consumption would therefore be expected to differ less across countries than the volume of government-provided services. Indeed, in spite of the fact that public-sector provision of human services is a larger share of GDP in western Europe than in the United States, 15.1 per cent as against 11.9 per cent, total (public plus privately provided) consumption of such services is larger in the United States than in Europe, 19.1 per cent as against 16.0 per cent (Table 1). In fact, this is the case for both education and health care – possibly partly reflecting a high income elasticity of demand for such services (with an 'automatic' supply response when such services are provided by markets).

It is probably easier to boost the aggregate consumption of human services by subsidies than by direct government provision – although the opposite is often asserted to be the case. (The government can be rather confident that general subsidies do increase the aggregate consumption of such services, in contrast to the case of direct government provision.) It is also *cheaper* for the government to boost such consumption by a certain volume by way of a subsidy than by way of direct provision. (While in the case of government provision the government has to finance the entire spending on such consumption, it has to finance only a fraction of total spending in the case of subsidies.)

There are other important differences between subsidies and direct government provision of human services. A subsidy allows the price to clear the market (zero excess demand), which implies that individuals will be able to choose volume and quality themselves, based on each individual's preferences and budget constraint. When judging the usefulness of allowing freedom

of choice in the consumption of human services, it is, however, important to consider a number of other aspects as well, such as the efficiency of production and the quality of the services, the distribution of the services among households, and possible tendencies towards clustering ('segregation') of specific types of consumers (in terms of income, education, ethnicity, ideology and so on) on specific providers.

The age-specific nature of public social spending, of course, results in redistribution of resources (income as well as human services) over each individual's life cycle (intra-individual redistribution). Usually, resources are transferred to individuals below age 20–25 and above age 60–65, and extracted (via taxes) from individuals in the age groups in-between. Indeed, we may regard public financing of education as a (collectively decided) loan from the middle-aged to the young, and public financing of pensions as a subsequent payback of the loan via payroll taxes (Becker and Murphy 1988). By these arrangements, two problems of intergenerational contracting are solved simultaneously: a liquidity constraint is removed for investment in human capital, and a universal pension system is created.

Indeed, in countries with highly universal welfare-state arrangements, the *bulk* of social spending constitutes such intra-individual redistribution rather than interindividual redistribution of lifetime income ('wealth'), in contrast to countries with strongly targeted systems. For instance, the universal character of public social spending in Sweden and Italy helps explain the high shares of aggregate social spending that constitute intra-individual redistribution over the individual's life cycle in these countries (83 per cent and 76 per cent, respectively, according to Finance Department, Sweden 2003; O'Donoghue 2001). The figure is, however, also boosted in countries, such as Sweden, where benefits usually are taxed. By contrast, the strongly targeted character of the social system in Australia helps explain its rather modest fraction of public social spending that consists of such intra-individual redistribution (38–52 per cent according to Falkingham and Harding 1996). As pointed out above, in countries with large intra-individual redistribution over each individual's life cycle, the remaining part of public social spending (and its financing) is often sufficient, however, to generate considerable inter-individual redistribution of yearly income.

So far we know very little about the consequences of welfare-state arrangements for the distribution of *lifetime* disposable income. However, some simulations based on Swedish data indicate that lifetime income ('wealth') is to a considerable extent redistributed from the upper part of the distribution of lifetime income (the highest two quintiles) to the lower part (the lowest three quintiles) – if we abstract from conceivable general equilibrium effects (Finance Department, Sweden 2003).

Problems

It is useful to classify major problems of contemporary welfare-state arrangements into (a) basically exogenous disturbances and (b) basically endogenous developments caused by the welfare state itself.

Exogenous Factors

It is a commonplace that recent and predicted future changes in *demography* in developed countries, in particular the 'graying' of the population, simultaneously boost social spending and have a negative influence on the tax base – since there are seldom automatic adjustments of social security contributions and benefit rules in response to changes in demography. Indeed, in the EU-19 the number of individuals above the statutory retirement age is already close to 25 per cent of the number of individuals of working age – and is projected to rise to about twice that figure or even more within three or four decades. It is difficult to alleviate this problem in the medium term except via immigration and tougher social-insurance legislation in the form of higher contribution rates, reduced benefits, stricter controls, and a higher effective retirement age.

The slowdown in the rate of *productivity growth* in the market sector in developed countries after the mid-1970s has created more or less

the same financing problems, since neither the contribution rates nor the benefit rules in the social insurance systems are automatically (fully) adjusted to changes in productivity growth. So far, politicians have usually tried to deal with this problem in the same way as they have tried to adapt to demographic changes, that is, by ad hoc reductions in benefits and increases in social-insurance contributions. In recent years, the *internationalization* (globalization) of national economies has become perhaps the most hotly debated exogenous factor behind actual and predicted future welfare-state problems in developed countries. International trade theory predicts that the entry into the world economy of a number of countries with abundant low-wage labour (including China, India, the former Soviet republics and countries in eastern Europe) will reduce both the wage-income share of national income and the relative wages of low-skilled workers. Clearly, these consequences are bound to create problems for policy ambitions concerning the distribution of income in many developed countries. It is often also argued that the rate of structural change is likely to accelerate, thereby resulting in tendencies toward higher structural unemployment, due primarily to limited flexibility of the allocative mechanisms in national economies. With given social legislation, this would certainly boost transfer payments (including unemployment benefits) and give rise to an erosion of the tax base, thus threatening the financial sustainability of the welfare state.

If such problems were actually to arise, the standard policy advice is, of course, measures to promote the flexibility of domestic product and factor markets, for instance, along the lines of the so-called Lisbon Agreement among EU countries in 2000. Important examples are retraining of workers, easier entry and expansion of firms, less strict job-security legislation, and more flexible relative wage rates – possibly combined with employment subsidies for low-skilled workers (the 'working poor').

Another common worry in connection with the internationalization process is that important tax bases tend to become more internationally mobile. While, so far, this has occurred mainly for capital income, there is a possibility that similar (although less pronounced) consequences will emerge for other tax bases as well, possibly resulting in increased tendencies towards tax competition among governments. To the extent that such developments actually occur, increased international tax coordination ('harmonization') is perhaps the most frequently recommended, and predicted, policy response.

Moreover, increased migration to developed countries may place an additional strain on the financial position of various welfare-state arrangements, in spite of the fact that such immigration is likely to 'improve' the age structure of the population, since migrants may face difficulties in obtaining employment. Poorly functioning labour markets, partly as a result of regulated wages, would be an explanation. To the extent that governments are unable to alleviate these deficiencies, politicians will most likely remain under political pressure to stiffen the restrictions on immigration.

Internationalization is, however, not the main reason for the serious *unemployment problems* in Western Europe in recent decades, boosting welfare-state spending and damaging the tax base. Regardless of whether the background is a higher equilibrium unemployment rate or increased unemployment persistence (after unemployment-creating macroeconomic shocks or increased microeconomic turmoil), approximately the same types of structural reform are potentially useful. If the problems are caused by persistence mechanisms, there is, however, also a strong case for liberalizing job-security legislation, and adopting other policy measures that reduce the market power of labour-market insiders – both phenomena contributing to inertia of the employment level. Counter-cyclical demand management policies (monetary and fiscal policy) are also more useful if it is unemployment persistence (after unemployment-creating macroeconomic shocks), rather than higher equilibrium unemployment, that is the problem.

Baumol's 'cost disease' (1967) regarding labour-intensive human services – such as childcare, education and old-age care – is another largely exogenous threat to the financial viability

of today's welfare-state arrangements. More specifically, since the relative costs of such services tend to increase over time (owing to slow productivity growth for such services), it will be necessary to raise tax rates gradually (without apparent limits) in countries where these services are tax-financed, even if the provision of such services is allowed to increase only rather slowly. The problem is somewhat different in the case of health care. After all, productivity in the health-care sector tends to rise rather rapidly along with advances in medicine and surgical techniques. However, since these improvements partly take the form of increased possibilities to treat health problems that could not be treated before, it is unavoidable that the demand for health care will also be boosted (at given incomes and prices). As a result, health care will, in fact, be exposed to similar financing problems as other human services, although partly for different reasons.

As a result of Baumol's cost disease, countries that today rely mainly on tax financing of human services will sooner or later have to limit the rate of expansion of such services (to the same rates as the increase in labour productivity of such services) or they will have to introduce complementary methods to finance human services, such as user fees and (voluntary or mandatory) insurance. Indeed, countries that are unwilling to accept such complementary financing methods may very well find themselves unable to finance equally large volumes of human services as countries with other financing methods. Perhaps these considerations help explain why both education and health spending, as mentioned above, are higher in the United States than in western European countries (although relatively high wages in the health sector in the United States is another explanation).

Endogenous Factors

In contrast to the welfare-state problems discussed above, disincentive effects via tax distortions and moral hazard are (by definition) the result of endogenous adjustments of individuals to the welfare-state itself. In the case of income insurance, moral hazard (*ex post)* arises simply because the individual will be able to choose more leisure at a very low cost to himself in terms of lost income. It is also well known that health-care insurance induces some patients to ask for excessive medical tests and expensive treatment, demands that many physicians may be willing to satisfy.

Formally, the individual will (tautologically) choose work rather than benefits only if $u[w(1-t)] > u(bw) + \alpha - f(n)$, where u is consumption utility, w the wage rate, t the average tax rate, b the benefit (replacement) rate, and a the *difference* between the utility of leisure and the intrinsic utility that one may derive from work. $f(n)$ denotes the disutility of stigmatization when breaking the prevailing work norm, where n is the aggregate number of individuals (or peers) who actually obey the work norm (or a norm against living on government benefits). I assume that the disutility of being stigmatized increases with the number of individuals who work rather than live on benefits; hence, $f'(n) > 0$. If we abstract, for the time being, from the social norms expressed by the stigmatization term $f(n)$, the individual may prefer to live on benefits rather than on work already when the after-tax rate $(1 - t)$ is only modestly higher than the benefit rate (b), provided he evaluates leisure at least somewhat more than work (so that a is at least somewhat positive).

Of course, sufficiently strong social norms in favour of work (or against living on benefits), that is, a sufficiently high value of the $f(\cdot)$ function may prevent widespread and frequent reliance on benefits even if the difference in income when one works and when one lives on benefits is quite small. After a while, however, some 'entrepreneurial' individuals may be tempted to exploit the benefit systems. As a result, social norms in favour of work (against exploiting the benefit systems) may erode among others as well (Lindbeck et al. 1999; Lindbeck and Persson 2006). The long-term negative effects of more generous welfare-state arrangements on aggregate labour supply may then be stronger than suggested by traditional microeconomic studies of the elasticity of labour supply with respect to after-tax wage rates. (Empirical research on the role of social norms in favour of work or against living on benefits is, however, still at an early stage.)

As an illustration of the potential importance of moral hazard for per capita hours of work, we may note that about a fifth of the population of working age (15–64) in western Europe today (2006) lives on various cash transfers from the government – the most important examples include unemployment benefits, labour-market programmes, social assistance, sick-pay insurance, and early retirement pensions (OECD 2003, pp. 188–90). Such moral hazard effects of generous welfare-state arrangements in western Europe are, therefore, an important explanation for the limited per capita hours of work in that part of the world. As a comparison, per capita hours of work (per year) in the United States are between 30 per cent and 50 per cent higher than in western Europe. (Prescott 2004, has instead tried to explain this phenomenon by the higher *marginal* tax rates in western Europe, assuming quite high labour-supply elasticities with respect to after-tax wage rates.)

The character and size of the incentive effects of welfare-state arrangements depend, of course, on the specific rules of both the benefit arrangements and the financing of these. For instance, to the extent that tax-financed benefits are paid to retired individuals rather than to individuals of working age, the negative substitution effects of the tax wedges on labour supply are counteracted by positive income effects of the tax payment (since, in this case, the taxpayers of working age do not get anything back in exchange for the tax payments). It is also well known that the negative substitution effects of marginal tax wedges on the labour supply are mitigated if there is a (positive) link between the individual's contributions to various social-insurance systems and his expected future benefits – as in the case of actuarially fair or 'quasi-actuarial' social-insurance arrangements – provided the individual is aware of this link. It is also a commonplace that negative incentives to acquire education as a result of marginal (in particular, progressive) tax rates are often counteracted, or perhaps even overcompensated, by subsidies to investment in human capital. Moreover, in some countries tax revenues are used to finance services that are close substitutes for home production, and hence complements to work in the open labour market. Subsidies to childcare and old-age care outside the family are important examples. In this special case, the negative substitution effects of tax wedges on the labour supply would be counteracted by positive cross-substitution effects on labour supply of the subsidized (or directly provided) services.

From an empirical point of view, the consequences of welfare-state spending on the efficiency and growth of the national economy are, of course, a perennial issue. In the case of countries with modest levels of such spending, economists generally agree that the positive effects of higher welfare-state spending on economic efficiency and economic growth are likely to dominate over the negative effects. This is particularly likely if increased public spending, starting from low levels, is concentrated on features such as sanitation, basic health care, elementary education and infrastructure, and if more comprehensive and generous income protection would further mitigate tendencies towards social unrest. However, there is also general agreement that, sooner or later, ever-increasing social spending will render the net effects on economic efficiency and growth negative, although it is difficult to identify the turning point.

The complexities of analysing and aggregating the effects of various types of benefit arrangements, and related taxes, have prompted many economists to try to find short cuts, by simply regressing either the level or (more often) the aggregate growth rate of per capita GDP on broad aggregates of taxes or government spending programmes. It is a fair summary of this huge literature that there is stronger support for the hypothesis that the effects of higher spending and taxes in today's developed countries are negative rather than positive. (Basically, studies conducted since around 1990 conclude that the effects are either negative or absent.) However, such aggregate studies suffer from well-known methodological problems.

New Requirements

The modern welfare state is a success in the sense that it has contributed to solving a number of

potentially serious social problems. It encounters, however, financial difficulties in several countries. Some welfare-state arrangements, and their financing, have also created *new* problems, including benefit dependency and other incentive effects. These developments are, of course, the background for ongoing and planned reforms of, and retreats from, existing welfare-state arrangements in a number of countries.

At the same time, strong political demands have emerged for new or improved social arrangements in several areas. For instance, increased female labour-force participation has raised the demand for paid parental leave, subsidized childcare, and old-age care outside the family – basically to facilitate everyday life among families with two income earners. In some countries, such arrangements are also regarded as important methods for restoring rapidly falling birth rates. The reduced stability of the family has also generated a political demand for legislated property rights in spouses' social-insurance benefits, in particular pensions.

There is also evidence of increasing individualization of values and lifestyle in developed countries, as compared with a number of decades ago, when today's welfare-state arrangements were designed (for evidence of such value changes, see Inglehart et al. 2004). Obvious ways of adjusting various benefit systems to these new values are more individually differentiated and portable social entitlements (nationally as well as internationally), as well as increased freedom for the individual to choose type of (mandatory) income insurance and quality of various types of (subsidized) human services, for instance via voucher systems (in a wide sense of the term).

Moreover, the incidence of economic and social misery among specific minority groups has recently increased in several developed countries – partly as a result of rising long-term unemployment, immigration of low-skilled groups from poor countries, alcoholism, drug abuse and de-institutionalization of the mentally ill – the 'truly disadvantaged' individuals. These problems require more than a generally improved situation in the labour market; new types of *targeting* social policies are necessary to help specific minority groups. A generally accepted view among social workers seems to be that it is also important to integrate more closely the administration of social insurance, social assistance, labour-market exchange systems, health care, rehabilitation, labour-market training and so on. Moreover, in some cases non-governmental organizations, including non-profit organizations, seem to be more successful than governmental organizations in such endeavours. These observations raise the issue of the potential usefulness of new divisions of tasks among governments, markets, the family, and civil society.

See Also

- ▶ Health Insurance, Economics of
- ▶ Labour Supply
- ▶ Social Insurance
- ▶ Social Norms
- ▶ Taxation of the Family

Bibliography

Adema, W., and M. Ladaique. 2005. Net social expenditure, 2005 edition: More comprehensive measures of social support. Social, employment and migration working papers no. 29. Paris: Directorate for Employment, Labour and Social Affairs, OECD.

Alesina, A., and D. Rodrik. 1994. Distributive politics and economic growth. *Quarterly Journal of Economics* 109: 465–490.

Atkinson, A. 1991. Social insurance: The fifteenth annual lecture of the Geneva association. *Geneva Papers on Risk and Insurance Theory* 16: 113–131.

Barr, N. 1998. *The economics of the welfare state*. 3rd ed. Oxford: Oxford University Press.

Baumol, W. 1967. Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review* 57: 415–426.

Becker, G., and K. Murphy. 1988. The family and the state. *Journal of Law and Economics* 31: 1–18.

Esping-Andersen, G. 1990. *The three worlds of welfare capitalism*. Cambridge: Polity Press.

Falkingham, J., and A. Harding. 1996. Poverty alleviation versus social insurance systems: A comparison of lifetime redistribution. Discussion paper no. 12. National Centre for Social and Economic Modelling, Faculty of Management, University of Canberra.

Finance Department, Sweden. 2003. *Fördelningen ur ett livscykelperspektiv [Distribution from a life-cycle perspective]*. *Långtidsutredningen*, bilaga 9. Stockholm: Finansdepartementet.

Flora, P., and J. Alber. 1981. Modernization, democratization, and the development of welfare states in Western Europe. In *The development of welfare states in Europe and America*, ed. P. Flora and A. Hedenheimer. New Brunswick: Transaction Press.

Forssell, Å., M. Medelberg, and A.-C. Ståhlberg. 2000. Olika transfereringssystem men lika inkomster. De äldres ekonomiska situation i ett internationellt perspektiv [Different transfer systems but equal income: the economic status of the elderly in an international perspective]. *Ekonomisk Debatt* 28: 143–158.

Inglehart, R., M. Basanez, J. Diez-Medrano, L. Halman, and R. Luijkx, eds. 2004. *Human beliefs and values: A cross-cultural sourcebook based on the 1999–2002 values surveys*. Mexico City: Siglo XXI Ed.

Kangas, O., and J. Palme. 1993. Statism eroded? Labor-market benefits and challenges to the Scandinavian welfare states. In *Welfare trends in the Scandinavian countries*, ed. E. Hansen et al. New York: M.E. Sharpe.

Korpi, W., and J. Palme. 1998. The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and poverty in the Western countries. *American Sociological Review* 63: 661–687.

Lindbeck, A., and M. Persson. 2006. A model of income insurance and social norms. Working paper no. 742. Stockholm: Institute for International Economic Studies.

Lindbeck, A., S. Nyberg, and J. Weibull. 1999. Social norms and economic incentives in the welfare state. *Quarterly Journal of Economics* 114: 1–35.

Lindert, P. 2004. *Growing public: Social spending and economic growth since the eighteenth century*. Cambridge, MA: Cambridge University Press.

O'Donoghue, C. 2001. Redistribution in the Irish tax-benefit system. Ph.D. thesis, London School of Economics.

OECD. 2004. *Education at a glance*. Paris: OECD.

OECD. 2005. *Society at a glance: OECD social indicators*. Paris: OECD.

OECD (Organisation for Economic Co-operation and Development). 2003. *OECD employment outlook*. Paris: OECD.

Pearson, M., and J. Martin. 2005. Should we extend the role of private social expenditure? Discussion paper no. 1544. Bonn: IZA.

Prescott, E. 2004. Why do Americans work so much more than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review* 28: 2–13.

Wells, David Ames (1828–1898)

Warren J. Samuels

Keywords

Free trade; Individualism; Laissez-faire; Wells, D. A.

JEL Classifications

B31

Wells was born on 17 June 1828 in Springfield, Massachusetts, and died on 5 November 1898 in Norwich, Connecticut. Trained at Williams College and Lawrence Scientific School at Harvard, Wells first taught and published as a geologist and chemist. After newspaper work, Wells turned to economics in his mid-forties. After publishing on the national debt, he was appointed to a series of federal and state tax positions, where he issued influential reports, revised tax laws, and originated the stamp system for collecting taxes on tobacco and liquor. He lectured in economics at Yale, Harvard and elsewhere, succeeded John Stuart Mill in 1874 as foreign associate of the French Academy, was president of the American Social Science Association, and received honorary degrees from Oxford, Harvard and Williams. His economic interests were practical and empirical, rather than theoretical; his place was transitional between the popular writer and the technically trained professional investigator.

Politically active, he was a leading exponent of laissez-faire doctrine, which he equated with individualism (in the manner of William Graham Sumner, with whom he was associated), free trade and the gold standard. Although an early protectionist disciple of Henry C. Carey, he later actively wrote and campaigned in favour of free trade. He opposed fiat money, the greenbacks and free silver. At one point he proposed the conversion of the greenbacks to interest-bearing government bonds; at another, he advocated a 'cremation theory of specie resumption', with the Secretary of the Treasury to burn a volume of greenbacks each day until they attained parity with gold.

Considered by some to be so doctrinaire as to be impervious to the stresses brought by industrialization in the late 19th century, he was none the less concerned with economic instability. Here he departed from orthodox doctrine, emphasizing the existence of unemployment due to both technology and overproduction relative to present demand, aggravated by the decline of available public lands as an alternative open to labour. His remedy was freer trade.

Through his will, he established the David Ames Wells Prizes in economics at Harvard University. Recipients have included John H. Williams, Jacob Viner, James W. Angell, Seymour E. Harris, Edward Chamberlin, Robert Triffin, Paul Samuelson, Lloyd Metzler, Robert Solow, Peter Kenen, Lester Thurow and Donald McCloskey.

Selected Works

1864. *Our burden and our strength.* New York: Loyal Publication Society.
1875. *The cremation theory of specie resumption.* New York: Putnam's.
1876. *Robinson Crusoe's money.* New York: P. Smith.
1877. *The silver question*. New York: Putnam's.
1884a. *A primer of tariff reform*. New York: New York State Revenue Reform League.
1884b. (With W.G. Sumner.) *Political economy and political science.* New York: Society for Political Education.
1885. *Practical economics*. New York: Putnam's.
1888. *Relation of the tariff to wages.* New York: Putnam's.
1889. *Recent economic changes*. New York: D. Appleton.
1900. *The theory and practice of taxation*. New York: D. Appleton.

Bibliography

Joyner, F.B. 1939. *David Ames Wells: Champion of free trade*. Cedar Rapids: Torch Press.

West, Edward (1782–1828)

Mark Blaug

Keywords

Differential rent; Diminishing returns; Free trade; Malthus, T. R.; Market price; Natural price; Ricardo, D.; Wages fund; West, E.

JEL Classifications

B31

Edward West is remembered – if he is remembered at all – for having stated the theory of differential rent based on the principle of diminishing returns in a long pamphlet just before Ricardo did so, and in virtually the same form and language. This has earned West the title of 'the first, though not the name-father and greatest of the "Ricardian" school' (Cannan 1893, p. 219). However, it appears that Ricardo developed the principle of diminishing returns independently of West and even of Malthus (who also published the idea more or less simultaneously) and at any rate Ricardo's exposition in his *Essay on Profits* (1815) was clearer then anyone else's, was more carefully set out and went beyond West in spelling out its implications for the distribution of income between wages, profits and rent. In addition to his *Essay on the Application of Capital to Land, with Observations Shewing the Impolicy of any Great Restriction of the Importation of Corn* (1815), West only wrote one other work on economics, a short book entitled *Price of Corn and Wages of Labour, with Observations upon Dr. Smith, S, Mr. Ricardo, S, and Mr. Malthus's Doctrines Upon those Subjects* (1826). At the time of his death, he was working on a treatise in political economy, the manuscript of which has been lost.

West was born in 1782 near London, educated at Harrow and University College, Oxford (where he studied classics and mathematics), and then went on to study law. In 1817, two years after the *Essay on the* Application of *Capital to Land,* he published a major treatise on the law of 'extents' (indemnities against direct or indirect debts to royalty), which was instrumental in reforming the use of extents in the Court of Chancery. In 1822 he was knighted and appointed Recorder of Bombay, followed two years later by the post of Chief Justice of the Crown in the Bombay province of India. The publication of his book on the *Price of Corn* in 1826 showed that he maintained his interests in economics until his death in India in 1828.

The similarity between the ways in which both West and Ricardo expressed the principle of

diminishing returns in agriculture – in terms of diminishing average rather than marginal products of composite doses of capital-and-labour applied to a fixed quantity of heterogeneous land and inclusive, not exclusive, of technical progress in agriculture – is startling, and so is the fact that both of them employed it to deduce a falling rate of profit on capital that could be postponed, but not permanently reversed, by the abolition of the Corn Laws. The only striking difference between the two 1815 pamphlets lay in the implications the two authors deduced from diminishing returns: Ricardo inferred that rents per acre would rise as more capital and labour were applied to ever inferior land, while West inferred that rents would fall, so that free trade would benefit landlords as well as capitalists and workers. This was a point on which West later changed his mind: in the *Price of Corn,* he agreed with Ricardo's inference about both rents per acre and the rental share. Unlike Ricardo, West realized that free trade would not imply complete specialization as between manufacturing in Britain and agriculture in Britain's trading partners: diminishing returns would operate abroad to raise the price of exported corn even as free trade would diminish the pressure on the costs of raising corn at home, so that eventually 'the actual price of both in the market must meet'. In this way, he met what was at the time a critical objection to the notion of free trade, namely that it would make Britain for ever dependent on foreign food supplies.

West's *Price of Corn* is a notable book if only because it was virtually the first work to attack the wages fund doctrine embedded in the writings of Adam Smith and Ricardo. 'The opinion that the demand for labour is regulated solely by the amount of capital', West asserted, 'has led perhaps to more false conclusions in the science than any other cause'. The demand for labour, he insisted, is not governed by the stock of wage goods inherited from the past but by the total level of private and public investment and consumption spending in the economy. It followed, he concluded, that 'the demand for the money wages of labour may be increased without any increase of the capital of the country'. The book contained a number of other insights, although opinions must differ as to how original these really were. There was the idea that price is determined by demand and supply, each of them considered as schedules of quantities at various hypothetical prices (an idea also found in Malthus); that the long-run 'natural' price of commodities is equal to average costs, including normal profits; that all manufacturing firms have identical cost functions; that the short-run market price of industrial goods cannot fall below average variable costs; and that the effect of a change in agricultural output on the price of corn depends on both the price and income elasticity of demand for corn. For some commentators these insights make him a 'Marshallian before Alfred Marshall' rather than a Ricardian before Ricardo (Grampp 1970).

Bibliography

Cannan, E. 1893. *A history of the theories of production and distribution in English political economy from 1776 to 1848.* London: Staples Press, 1953.

Grampp, W.D. 1970. Edward West reconsidered. *History of Political Economy* 2: 316–343.

Westergaard, Harald Ludvig (1853–1936)

Hans Brems

Westergaard was born on 19 April 1853 in Copenhagen. At the University of Copenhagen he took degrees in mathematics (1874) and economics (1877). He studied in Britain and Germany (1877–8), taught statistics and economics at his alma mater from 1883 to 1924, and died on 13 December 1936 in Copenhagen.

As an economist, Westergaard rejected English classical theory on two grounds. First, his warm heart reacted against a dismissal of social reform as futile. Like his German historical colleagues, Westergaard became an early champion of *Sozialpolitik*. Second, his keen mind reacted against a dismissal of the demand side. In his

Indledning (1891) he introduced the post- 1870 revolution of economic theory in the form received from his friend Jevons. His first article (1876) had accepted cardinal utility to the point of making interpersonal utility comparisons: equalizing the distribution of wealth would increase community utility. Westergaard urged Jevons (1879: preface) to derive the Jevonian marginal equalities simply as the first-order conditions for an individual utility maximum, but Jevons was not up to it.

Jevons never reached a fully general economic equilibrium: his households were do-ityourself households engaged in barter. By contrast, Walras separated households from industry, found physical quantities of outputs supplied by industry and demanded by households, physical quantities of inputs supplied by households and demanded by industry, and relative prices at which all such outputs and inputs were transacted. By introducing the post-1870 revolution in its Walrasian form, Wicksell and Cassel gave Sweden a clear headstart. After Westergaard, the first Danish generation (Birck) remained Jevonian, and it took yet another generation (Zeuthen) for Walras to reach Denmark.

As a statistician, Westergaard did his most original work. His *Lehre* (1882) and his *Grundzüge* (1890) enjoyed international fame. A 50-page English-language summary, 'Scope' (1916), was published by the American Statistical Association. Westergaard's tools were the binomial and normal distributions combined with a deep analysis of the data divided into rational subgroups. He used his tools to the hilt on demographic and anthropometric data but was always less interested in mathematical rigour than in finding what was hiding behind his data and his errors of measurement. In the fine art of listening to the voice of numbers, few practitioners have had an ear as sensitive as his.

Westergaard's normal distribution was at home in demography and anthropometry, and he was always reluctant to move on to less friendly habitats. Although trained in the triad, then unique, of economic theory, mathematics and statistics, he never became an econometrician: to let economic theory suggest possible forms of a regression equation, to estimate the parameters of each form, and to let correlation analysis decide the choice among forms seemed frivolous to him. Eight years ahead of Moore, Westergaard's doctoral candidate Mackeprang (1906) estimated demand functions for 24 commodities. Westergaard was unimpressed and in his farewell lecture in 1924 still considered correlation analysis a fad.

Out of sympathy with the trends of his chosen field, Westergaard turned to its history and gave us his only English-language full-length book, his *Contributions* (1932). Wide coverage, lucid restatement, and historical accuracy made it a classic.

Selected Works

1876. Den moralske Formue og det moralske Haab (Moral fortune and moral expectation). *Tidsskrift for Matematik*.
1882. *Die Lehre von der Mortalität und Morbilität*. Jena: Fischer. 2nd enlarged ed, 1901.
1890. *Die Grundzüge der Theorie der Statistik*. Jena: Fischer. 2nd much enlarged ed, 1928.
1891. *Indledning til Studiet af Nationalkonomien* (Introduction to the study of political economy). Copenhagen: Philipsen.
1916. Scope and method of statistics. *Quarterly Publications of the American Statistical Association* 15.
1932. *Contributions to the history of statistics*. London: King.

Bibliography

Jevons, W.S. 1879. *The theory of political economy*, 2nd ed. London: Macmillan.
Kaergård, N. 1983. Marginalismens gennembrud i Danmark og mændene bag (The breakthrough of marginalism in Denmark and the men behind it). *Nationaløkonomisk Tidsskrift* 1.
Kaergård, N. 1984. The earliest history of econometrics: some neglected Danish contributions. *History of Political Economy* 16(3): 437–444.
Mackeprang, E. 1906. *Pristeorier (Price theories)*. Copenhagen: Bagge.

Rander Buch, K. 1968. Harald Westergaard. In *International encyclopedia of the social sciences*, vol. 16, 528–529. New York: Macmillan.
Royal Statistical Society (anonymous). 1937. Obituary. Harald Westergaard, 1853–1937. *Journal of the Royal Statistical Society* 100: 149–150.

Whately, Richard (1787–1863)

R. D. Collison Black

Whately was Fellow of Oriel College, Oxford, 1811–22; rector of Halesworth, Suffolk, 1822–5; principal of St Alban's Hall, 1825–31; Drummond Professor of Political Economy in the University of Oxford, 1829–31; and Archbishop of Dublin, 1831–63.

In the one course of lectures on political economy which he published, Whately displayed flashes of originality, as in his famous aphorism on the labour theory of value, 'It is not that pearls fetch a high price *because* men have dived for them, but on the contrary men dive for them because they fetch a high price', and in his suggestion to re-name the subject 'Catallactics, or the science of exchanges' (1832, pp. 6 and 253). In later years his *Easy Lessons on Money Matters*, the first attempt to present classical political economy in a primary school book, went through 16 editions and was translated into at least three languages.

Although Whately's period as a professor of political economy was only a brief interlude in a long career as theologian and churchman, he exerted a continuing influence on the subject in a variety of ways apart from his writings. He was a lifelong friend of Nassau Senior, his Oxford contemporary and predecessor in the Drummond chair; each influenced the other's thinking on matters of both economic theory and policy.

After his move to Dublin, Whately founded and personally financed the chair of political economy in Trinity College, basing it on the Oxford model and examining the candidates himself. A liberal on matters of social policy, he opposed the introduction of the workhouse system in to Ireland, urging wider measures of economic development.

Selected Works

1831. *Introductory lectures on political economy.* London: B. Fellowes. 2nd edn (including an additional lecture), 1832.
1833. *Easy lessons on money matters, for the use of young people*. London: J.W. Parker for the Society for the Promotion of Christian Knowledge.

Bibliography

Goldstrom, J.M. 1966. Richard Whately and political economy in school books, 1833–80. *Irish Historical Studies* 15: 131–146.
Whately, E.J. 1866. Life and correspondence of Richard Whately, 2 vols. London: Longmans, Green, and Co.

Wheatley, John (1772–1830)

Frank Whitson Fetter

Keywords
Bank Charter Act (1844); Note issue; Wheatley, J.

JEL Classifications
B31

Lawyer and economist, Wheatley was born in Erith, Kent, of a prominent landed and military family, and died at sea on a voyage from South Africa to England. A memorial plaque to him is in the Wheatley chapel of the Erith parish church. At Oxford he was a member of Christ Church, and after receiving his BA in 1793 was admitted to Lincoln's Inn, but his activity in the law was limited, and his life was devoted largely to writing on economics and playing a small part

in Whig politics. With him at Christ Church was Charles Watkin Williams Wynn, nephew of Lord Grenville, and Wheatley was active in support of Grenville's successful campaign in 1809 for Chancellor of Oxford University; he had correspondence with Wynn in 1812 about running for Parliament on the Whig ticket, but nothing came of this; a book of 1816 took the form of a letter to Lord Grenville, and his pamphlet of 1823 was a letter to Wynn.

Wheatley published ten books and brochures, two of these in India and one in South Africa. He lived in these two countries for the last nine years of his life, evidently to escape creditors. His works published in India and South Africa received little contemporary attention, and today are extremely rare. Of the others, *Remarks on Currency and Commerce* (1803), and the first volume of *An Essay on the Theory of Money and Principles of Commerce* (1807) received the most contemporary attention, and best stated his theoretical position on the monetary controversy that followed the suspension of cash payments by the Bank of England in 1797. Wheatley stated, in an even more extreme way than Ricardo did later, that exchange fluctuations were due exclusively to domestic price changes, and that the Bank of England, through its credit policy, could control prices, and thus exchange rates. These books of 1803 and 1807 criticized the Bank for its monetary expansion, but following the resumption of specie payments in 1821 Wheatley in his book of 1822 had become a severe critic of both the Bank and of the Tory government for the price deflation. His efforts, both in his book and in correspondence with Whig leaders, to launch an attack on the government's monetary policy, made no headway. In several publications he stressed the danger, for monetary stability, of permitting the issue of notes by banks other than the Bank of England. No economist of his period so well anticipated the note issue provisions of the Bank Act of 1844, which led to the elimination of all notes other than of the Bank of England.

Wheatley's views on political issues were something of a paradox. As a Whig he was frequently a voice for reform; with the background of the landed gentry he sometimes disagreed with Whig positions that threatened the supremacy of the landed aristocracy. He supported free trade, a commercial union with France, removal of restrictions on West Indian trade, and the abolition of slavery. He favoured primogeniture, maintenance of great landed estates, the political supremacy of the landed gentry, and an unreformed Parliament. In foreign policy his imperialist views foreshadowed the idea of the 'white man's burden'.

Selected Works

1803. *Remarks on currency and commerce.* London.
1805. *Thoughts on the object of a foreign subsidy.* London.
1807, 1822. *An essay on the theory of money and principles of commerce,* vol. 1. London, 1807; vol. 2, London, 1822.
1807. *A letter to Lord Grenville, on the distress of the country.* London.
1819. *A Report on the reports of the bank committee.* Shrewsbury.
1821. *A plan to relieve the country from its difficulties.* Shrewsbury. This short pamphlet is an extract from the book of 1822 that appeared shortly afterwards.
1823. *Letter to the Rt. Hon. Charles Watkin Williams Wynn, President of the Board of Control, on the Latent Resources of India.* Calcutta.
1824. *A Letter to his Grace the Duke of Devonshire on the State of Ireland, and on the General Effects of Colonization.* Calcutta.
1828. *Tempora praeterita: Or, more currency and more corn.* Cape Town. This was published anonymously, but in correspondence Wheatley admitted authorship.

The Wheatley letters to Lord Grey are in the Grey of Howick papers at the University of Durham; the Wheatley letters to Charles Watkin Williams Wynn are at the National Library of Wales at Aberystwyth.

Bibliography

Fetter, F.W. 1942. The life and writings of John Wheatley. *Journal of Political Economy* 50: 357–376.
Viner, J. 1937. *Studies in the theory of international trade*. New York: Harper & Brothers.

Whewell, William (1799–1866)

G. Campanelli

Keywords

Fixed capital; Mathematical economics; Mathematical method in political economy; Production price; Whewell, W.

JEL Classifications

B31

Whewell was born in Lancaster and died in Cambridge. He received his early education at Lancaster Grammar School and Heversham School, Westmoreland, and in 1812 he went up to Trinity College, Cambridge. In 1817 he became a Fellow of the college, in 1823 a tutor. In 1841 he was made Master, an appointment which he held until his death.

Whewell was at the centre of a 'network' of Cambridge scientists and exercised considerable influence upon scientific and philosophic circles in Victorian England. In 1820 he became a Lecturer in Mathematics, in 1828 he was appointed Professor of Mineralogy and in 1838 Professor of Moral Philosophy. He was active as an honorary member of 25 scientific, historical and philosophical societies in several countries. To mention a few of the most important in England: he was one of the founders of the Cambridge Philosophical Society in 1818; in 1820 he was elected a Fellow of the Royal Society; in 1831 he became a member of the British Association and in 1841 was appointed President.

Whewell was primarily a philosopher and mathematician, and he published his major works in these fields (Whewell 1837, 1840). Political economy was one of the many other subjects dealt with by him. However, his contributions in this field – written over the whole period from 1829 to 1862 – give clear proof that his interest in economics was lifelong. Whewell's major works in political economy were four papers on mathematical economics which were read before the Cambridge Philosophical Society (Whewell 1829, 1831, 1850a, b) and a book – *Six Lectures on Political Economy* (1862) – which was composed for the edification of the Prince of Wales, the future Edward VII. In the *Six Lectures* Whewell presented, in a very elementary way, the principal ideas of Smith, Ricardo and Jones.

Whewell's four papers represent the earliest *systematic* application of mathematical symbols of political economy in England. Whewell believed that the arithmetic used by classical economists was inadequate, and that the more general language of algebra should take its place. He pointed out that the adoption of a mathematical method had two main advantages. Firstly, that many aspects of political economy could be presented in a more simple, clear and systematic form. Secondly – and more importantly – the use of mathematics could help to avoid the danger of drawing false conclusions for assumptions made. To illustrate this point, in his 1829 paper Whewell used mathematics to discuss Ricardo's theory of the incidence of a tax on wages. Ricardo had argued – against Smith – that a rise in the prices of goods due to a rise in wages would in turn affect wages and 'the action and reaction first of wages on goods and then of goods on wages, will be extended without any assignable limits' (Ricardo 1821, p. 225). Whewell, on the contrary, showed that if Ricardo had considered the mathematical implications of his theory, he would have found that an unlimited rise in prices and wages was impossible. Indeed, if it is assumed that only a part of the value of goods is wages, and only a part of the labourer's consumption consists of manufactured goods, then the paths that both prices and wages follow take the form of geometric series which converge.

But Whewell's most notable contribution to political economy was his mathematical formulation of Ricardo's theory, and in particular his analysis of fixed capital (Whewell 1831). This analysis is important not only because it represents the first mathematical treatment of machinery in Ricardo's model, but also and mainly because it constitutes an original contribution to the subject. In 1831 Whewell had already

provided an exact formulation for the reduction of fixed capital to dated quantities of labour. He also worked out a simple model to quantify the substitution effect between labour and machinery. Finally, through the annuity formula, he arrived at the equation which defines the production price in the presence of fixed capital.

These results also suggest that the dating of the genesis of fixed capital models may need to be reappraised, for it is usually thought that Bortkiewicz (1907) – on the basis of Dmitriev's contribution (1904) – was the first economist to treat fixed capital mathematically within the theory of production price.

Whewell has been consistently neglected in the history of economic analysis. The few authors who were acquainted with Whewell's work – of whom the most authoritative were Jevons (1871) and Schumpeter (1954) – considered his analysis as purely derivative: supposedly he merely translated into algebraic form results which others had previously expressed in non-mathematical language. The only exception was Walras, who regarded Whewell's contribution as 'really remarkable' (Walras 1875, p. 32). Whewell was in fact more than a translator: he was a major contributor to the early development of mathematical economics in England, and above all a pioneer in the general debate on fixed capital.

Selected Works

1829. *Mathematical exposition of some doctrines of political economy.* Cambridge: Cambridge Philosophical Society, 1830, Transaction 3.
1831. *Mathematical exposition of some of the leading doctrines of political economy: Second memoir.* Cambridge: Cambridge Philosophical Society, 1833, Transaction 4.
1837. *History of the inductive sciences, from the earliest to the present time,* 3rd ed. New York: Appleton, 1890.
1840. *The philosophy of the inductive sciences, founded upon their history,* 2nd ed. London: Parker, 1847.
1850a. *Mathematical exposition of some doctrines of political economy: Second memoir.* Cambridge: Cambridge Philosophical Society, 1856, Transaction 9, Part I.
1850b. *Mathematical exposition of certain doctrines of political economy: Third memoir.* Cambridge: Cambridge Philosophical Society, 1856, Transaction 9, Part II.
1862. *Six lectures on political economy.* Cambridge: Cambridge University Press.
1971. *Mathematical exposition of certain doctrines of political economy.* New York: A.-M. Kelley. A reprint of Whewell's four papers on mathematical economics (1829, 1831, 1850a and 1850b).

Bibliography

Bortkiewicz, L. 1907. Wertrechnung und Preisrechnung im Marxschen System. In *Archiv für Sozialwissenschaft und Sozialpolitik,* vol. 25. The second and the third parts are translated into English as 'Value and price in the Marxian system'. *International Economic Papers* No. 2: 5–60 (1952).
Dmitriev, V.K. 1904. In *Economic essays on value, competition and utility*, ed. D.M. Nuti. Cambridge: Cambridge University Press, 1974.
Jevons, W.S. 1871. *The theory of political economy.* London/New York: Macmillan.
Ricardo, D. 1821. In *On the principles of political economy and taxation*, ed. P. Sraffa. Cambridge: Cambridge University Press, 1951.
Schumpeter, J.A. 1954. *History of economic analysis*. New York: Oxford University Press.
Walras, L. 1875. Une branche nouvelle de la mathématique: de l'application des mathématiques à l'économie politique. Italian Trans. as 'Un nuovo ramo della matematica. Dell'applicazione della matematiche alla economia politica'. *Giornale degli Economisti* 3 (1876).

White, Harry Dexter (1892–1948)

James M. Boughton

Abstract

Harry Dexter White (1892–1948) was an American economist who spent most of his career at the US Treasury and the International Monetary Fund (IMF). Although he published

very little, he wrote a large number of internal memoranda that reveal an innovative and practical approach to economic policy. White was a New Deal liberal whose economic philosophy spanned the late classical and early Keynesian eras. His major achievement was his design for the IMF, the institution that he helped bring into existence in 1946.

Keywords

Bretton Woods system; International monetary fund; International monetary institutions; Taussig, F. W.; White, H. D

JEL Classifications

B31; F33; F55

Harry Dexter White, one of the most important government economists of the 20th century, is honoured primarily for his design for the International Monetary Fund (IMF). During the Second World War, after more than two years of negotiation with John Maynard Keynes, who had drafted a rival plan for a similar institution, White prevailed in most of the crucial facets of the design for the IMF. The charter for the new multilateral agency (known as the Articles of Agreement) was approved by 44 national delegations at Bretton Woods, New Hampshire, on 22 July 1994. The institution came into being on 27 December 1945, when the Articles were signed by representatives of 29 countries. While the IMF eventually became famous as the centrepiece of the post-war international financial system, White and his intellectual contributions remained largely obscured.

Harry White – like many Jewish boys of his generation, he added his middle name later – was born in Boston, Massachusetts on 9 October 1892. His parents, Isaac and Sarah White, were immigrants from Lithuania, which was then part of the Russian empire. Isaac White owned a small chain of hardware stores, where Harry worked for several years as a young man before joining the US army and serving in France during the First World War. Only after the war did he begin his university education, first at Columbia University in New York City and then at Stanford University in Palo Alto, California.

White received a BA in economics from Stanford in 1924 and an MA from the same university a year later. He then enrolled in the doctoral programme at Harvard, and in 1926 began a six-year stint as a Harvard instructor in economics while he completed his studies. He was awarded a Ph.D. in 1930, and in 1932 he took what would turn out to be his last academic post, as assistant professor at little-known Lawrence College in Appleton, Wisconsin. After one year he was promoted to professor, but the onset of the Great Depression and the election of Franklin Roosevelt as US President created the opportunity for White to begin his real life's work in Washington.

In 1934, Jacob Viner was assembling a 'brain trust' of economists at the US Treasury to implement Roosevelt's New Deal policies. Viner, who knew White through their mutual teacher Frank Taussig, invited White to join him for a summer doing economic research at the Treasury. White eagerly accepted, and he spent the next few months analysing the monetary consequences of the Gold Reserve Act, which Congress had passed that January. The summer assignment then was converted into a permanent position. White spent the next 12 years at the Treasury, through the Depression and the Second World War, rising eventually to the rank of assistant secretary: in effect, the chief economist of the department, reporting directly to secretary Henry Morgenthau, Jr.

In 1946, President Harry Truman named White to serve as the first US executive director at the IMF, the second most powerful position in the new institution. Within a year, after suffering a heart attack while on a Fund mission to South America, White resigned his post. He subsequently worked sporadically as a freelance consultant, primarily for the government of Mexico. On 16 August 1948, at his summer home in Peterborough, New Hampshire, he was stricken fatally with a second heart attack.

White's most important formative intellectual influence was Frank Taussig. Under Taussig's direction, White wrote his Ph.D. dissertation on the French international accounts, 1880–1913.

The thesis demonstrated that – contrary to the received wisdom of the time – France had not benefited economically from its considerable foreign financial investments. White also helped Taussig with the preparation of the third edition of the latter's classic analysis of tariffs. Most notably, he drafted a new section analysing the ability of several industries in the United States to withstand international competition without tariff protection. On the basis of these contributions, June Flanders categorized White as a 'late classical' economist along with Taussig, Viner and others.

Even before leaving Harvard in 1932, White began developing his thinking beyond this classical training into a more eclectic economic philosophy which prefigured Keynesianism. The clearest early indication of this development is an unpublished 1932 memorandum that White wrote jointly with Lauchlin Currie and P.T. Ellsworth, calling for a large-scale public works programme to combat the incipient depression, financed partly by issuance of Treasury debt and partly by open-market purchases by the Federal Reserve System. Throughout White's subsequent career, his prolific writing on macroeconomic policy – mostly in the form of internal Treasury memoranda – adhered closely to Keynesian analysis. On international policy, however, he held true to his Taussig training by consistently advocating openness to foreign competition and to multilateral finance. Although he concluded that controls on the international flow of financial capital had a place in the policy arsenal, he argued that they should be kept in reserve and used only in a crisis.

White's role in the Treasury placed him in position to influence US economic policy on many fronts, with particular emphasis on domestic and international monetary policies. One topic that occupied him throughout these 12 years was the role of gold in the monetary system. By Executive Order in 1933, Roosevelt had prohibited US citizens from owning monetary gold. The Gold Reserve Act had devalued the dollar in relation to gold, consolidated control of the official stock in the Treasury, and established the Exchange Stabilization Fund (ESF) as a vehicle for the Treasury to stabilize the value of the dollar in foreign exchange. Elsewhere, national governments were managing their monetary systems in a variety of ways, ranging from full adherence to the gold standard to free floating with a managed domestic monetary policy. This patchwork system was evidently not sustainable in the long run, but the alternatives – other than an unlikely return to the pre-war international gold standard – were not well understood. In a series of unpublished papers written over the course of his Treasury career, White concluded that gold should remain the underpinning of the US monetary system, but with escape valves that would enable the Treasury and the Federal Reserve to loosen the ties to gold in extreme circumstances. A close reading of these papers shows that White's analysis foreshadowed the system of 'fixed but adjustable' exchange rates that was the central feature of the post-war Bretton Woods system and was a precursor of the statecontingent policy rules that economists began modelling in the 1980s.

Early in his Treasury career, White realized that stabilizing the foreign exchange value of the dollar would require providing assistance to deficit countries, not just intervention in exchange markets. He therefore advocated expanding the range of uses of the resources held in the ESF. In January 1936, he organized an innovative repurchase agreement with the government of Mexico that set a precedent for later similar arrangements. The US Treasury provided dollars from the ESF to Mexico and received pesos in exchange, with silver pledged as collateral until the exchange was unwound (with interest) at a fixed subsequent date. Six years later, when White was developing his plan for the IMF, he built a similar currency-exchange mechanism into the design. That avoided the need for formal loan contracts between the agency and its member countries, and established the IMF as a unique financial institution through which countries could gain access to its resources without appearing to be in a supplicant borrower position. Eventually, the idea of the IMF as a lender would become more widely accepted, but at the outset this technical feature was a crucial element in enabling the White Plan to be politically acceptable.

When the Second World War broke out in Europe, White became alarmed over German influence in Latin America. He saw the region as of strategic importance to the United States and vulnerable to political unrest owing to pervasive poverty and monetary instability. To counteract the dangers, in 1940 he proposed the setting up of a regional bank through which the United States could offer loans for development and stabilization. While that proposal was under consideration, he continued to pursue bilateral deals to support friendly governments throughout Latin America. When the regional initiative failed to gain acceptance, he began sketching out an even more ambitious plan for a global institution – also to be financed primarily by the United States – to foster multilateral payments settlements, currency convertibility and monetary cooperation. The US entry into the war in December 1941 elevated the urgency of planning for post-war recovery, and Morgenthau immediately put White in charge of all international economic issues.

White circulated his first plan for the IMF in April 1942, shortly after Keynes developed an independent plan for an International Currency Union (ICU). The Keynes plan, based on the British system of bank overdrafts as a lending mechanism, would have created the ICU as a world central bank with powers to lend by issuing its own currency. White envisaged an institution with much more limited powers and resources. White and Keynes exchanged drafts, met several times during the war, and eventually agreed on the main elements for what would become the IMF Articles of Agreement. In 1944 and 1945, White devoted his energies almost entirely to negotiating the final design, first with Keynes and then with delegations from the other allied countries; then to generating a consensus at Bretton Woods; and finally to selling the proposal to the US Congress, which had to ratify the Articles before the IMF could come into existence.

White's liberal politics, his friendship and association with men who were or had been Communists, and his regular contacts with Russian economists in conjunction with the preparation and conduct of the Bretton Woods conference brought him under suspicion when the Cold War erupted and fear of closeted Communists gripped the United States. The FBI investigated him for three years, in the belief that those contacts must have constituted spying. In August 1948, three days before his death, White vigorously defended himself before the House Un-American Activities Committee of the US Congress. In 1953, at the height of McCarthyism, the administration of President Dwight Eisenhower dragged White's name back into the public consciousness by accusing Truman of having appointed White to his post at the IMF despite 'knowing' that White was a Soviet spy. Truman of course denied the charge, but the controversy continued through the ensuing decades.

See Also

- ▶ Bretton Woods System
- ▶ International Monetary Fund
- ▶ International Monetary Institutions
- ▶ Taussig, Frank William (1859–1940)

Bibliography

Boughton, J.M. 2002. Why White, not Keynes? Inventing the post-war international monetary system. In *The open-economy macro-model: Past, present and future*, ed. A. Arnon and W. Young. Boston: Kluwer Academic.

Boughton, J.M. 2004. New light on Harry Dexter White. *Journal of the History of Economic Thought* 26: 179–195.

Boughton, J.M., and R.J. Sandilands. 2003. Politics and the attack on FDR's economists: From the grand alliance to the cold war. *Intelligence and National Security* 18(3): 73–99.

Craig, B. 2004. *Treasonable doubt: The Harry Dexter White spy case*. Lawrence: University of Kansas Press.

Flanders, M.J. 1990. *International monetary economics, 1870–1960*. Cambridge: Cambridge University Press.

Laidler, D.E.W., and R.J. Sandilands. 2002. An early Harvard memorandum on antidepression policies: An introductory note. *History of Political Economy* 34(3): 515–532.

Rees, D. 1973. *Harry Dexter White: A study in paradox*. New York: Coward, McCann & Geoghegan.

Taussig, F.W. 1931. *Some aspects of the tariff question: An examination of the development of American industries under protection. Third enlarged edition, continued to 1930, with the cooperation of H. D. White*. Cambridge, MA: Harvard University Press.

White, H.D. 1933. *The French international accounts, 1880–1913*. Cambridge, MA: Harvard University Press.

Wicksell and Neoclassical Economics

Paul A. Samuelson

Knut Wicksell was long known as Scandinavia's Alfred Marshall, the leading economist of that region, whose microeconomics married Böm-Bawerk's time-phased interest theory with Walras's mathematical general equilibrium. His macroeconomics was thought to foreshadow Keynes's 1936 *General Theory*, even though it emphasized that the discrepancy between investment and saving is the cause merely of an inflationary trend in the general price level.

Usually when a great economist is translated into English, reputation is deflated. Not so with Wicksell: the 1930s appearance of his 1898 *interest and Prices* (Wicksell 1936) and of his *Lectures on Political Economy*, Volume I ('micro') and Volume II (money), which appeared in Swedish and German editions between 1901 and 1928 and was translated into English in 1934 (Wicksell 1934), sent his reputation soaring over his neoclassical contemporaries, Alfred Marshall in England and his great Swedish rival Gustav Cassel. Just below the pure-theory throne of Léon Walras sits Knut Wicksell: anyone who reads Hicks's seminal *Theory of Wages* (1932) will realize that Wicksell beings analysis to bear on the recurrent problems of out own age: as an example, his explication of how technical innovation can affect the distribution of income and the real-wage level is a quantum leap in sophistication over Ricardian and Marxian paradigms.

A decade after Wicksell's death in 1926, it was his saving–investment macroeconomics that economists most prized. All Scandinavians were neo-Wicksellians in the interval between Keynes's 1930 *Treatise on Monday* and *The General Theory of Employment, Interest and Money* (1936): on the continent, Austrians such as Ludwig von Mises, Freidrich von Hayek, and Gottfried von Haberler thought in the same mode; in America, the young Alvin Hansen was a Wicksellian fellow traveller in the period before his trek on the road to the Damascus of Keynes's *General Theory*. Although neo-Wicksellianism did weaken faith in Say's Laws, not much of it lives on in present-day economics. Now we realize that Gunnar Myrdal and Erik Lindahl were not anticipating in their Wicksellian debates of the early 1930s any general theory of *output as a whole*. Like the *Treatise on Monday* and Dennis Robertson's monetary paradigms, their primary focus was on the *price* level as the macrovariable of their equilibrium theory. True, Bertil Ohlin and Ragnar Frisch wrote presciently about induced changes in aggregate output as the Great Depression deepened and as efficacious fiscal and monetary policies were advocated. And in this they were mindful of Wicksell's grapplings with macroeconomics. But, as Don Patinkin (1982) has documented, these sage writers articulated no formal paradigm of effective demand comparable to that of the 1936 *General Theory* (or of Michal Kalecki's concomitant partial formulation of an aggregate–output model).

Before concentrating on the Wicksellian microeconomics that moderns see as the jewel in his crown, I should devote a few words to Wicksell's macroeconomics. Wicksell affirmed, rather than denied, that germ of truth in the Quantity Theory of Money and Prices which holds that raising in balance *all* nominal prices of goods, services and assets can leave all *real* supplies and demands and all relative prices intact. Actually, Wicksell's own theory of the business cycle was not a saving–investment analysis but rather an exogenous-shock theory emphasizing the innovations and technical changes that were emphasized by Joseph Schumpeter, Arthur Spiethoff, Gustav Cassel, and by the young Dennis Robertson and young Alvin Hansen. Wicksell's image of a rocking horse, which can be set into quasi-periodic motions even by random hammer blows, was later revived to good effect by Ragner Frisch and Jan Tinbergen.

Wicksell's saving and investment paradigm was essentially a theory of how the total supply of money will be driven secularly upward or downward by a Central Bank that insists on setting the market interest rate persistently below the *real* (or 'natural') interest rate defined neoclassically by the system's time-phased

technology and time-phased consuming preferences. if M grows secularly at 5 per cent per year above the trend growth rate of output – because of persistent perverse pegging of the market interest rate too low – then P in aggregate PQ grown at about 5 per cent with Q limited (over the business cycle) to about its same approximation to high-employment potential-output. In opposition to Wicksell's insistence on the goal of stability for the price level, his great Swedish contemporary David Davidson espoused a price level that fell proportionately to society's gain in productivity, an arguable thesis if we put aside pragmatic frictions. The Wicksell who uncharacteristically made a fetish of honouring prewar 1914 price levels could only with poor consistency make light of real-world frictions.

Wicksell's device of an interest rate specified by the bank-credit system was perhaps 'too clever by 'arf' and did mischief later in delaying decay of the Model T Keynesianism that dogmatically downplayed the potency of M-changes to affect real Q. The middle-aged Bourbons who fabricated the Radcliffe Committe Report in Britain (Radcliffe 1959), with its antiquated refusal to forget about great-depression liquidity traps, had their minds frozen in their salad days of 1936–39 along Wicksellian modes of though in which halvings in the general price and wage level merely halved the total money supply – and in which singular versions of the Keynesian systems were invoked, implicitly or explicitly, that hypothesized the near-vanishing of $\partial(\mathrm{MV})/\partial\mathrm{M}$.

For a neoclassical economist, the time to be born was by 1840, the epoch of Jevons, Walras and Menger. Vilfredo Pareto, Marshall, Wicksell and Cassel came too late for the feast – to say nothing of A.C. Pigou, Frank knight and Jacob Viner. Besides, Wicksell was a late bloomer, whose degree in mathematics was followed by a bohemian existence of preoccupation with anti-religion, anti-sexual puritanism, anti-alcoholism, anti-monarchism and anti-militarism. Just as birth control was beginning to catch hold in Europe, Wicksell was obsessed with Malthusian overpopulation and the law of diminishing returns.

By good luck and genius, this self-taught and unemployed post-Doc wrote in his forties a splendid synthesis of Böhm-Bawerk's capital theory with Walras's general equilibrium. See *Über Wert, Kapital und Rente* (1893a), translated as *Value, Capital and Rent* (1954), in which Wicksell builds a Jevons–Böhm model where output is increased when the time interval is enlarged between application of inputs and harvest of output. This marginal-productivity-of-time paradigm for a positive interest rate is second in importance only to Irving Fisher's 20th-century general equilibrium reformulation of Böhm-Bawerk's insights.

Only for singularly special technologies is it true that the interest rate equals the derivative of the value of total output with respect to the value of total capital, $\partial(\sum \mathrm{PQ})/\partial(\sum \mathrm{PK})$ – as Wicksell pointed out with reference to the technology of maturing wine. (His accusation (Wicksell 1954, pp. 141–2) that Thünen erred on this point is refuted in Samuelson's (1983, Equation A11) demonstration.) Such a discrepancy between the interest rate and this derivative is called a 'Wicksell effect' in the modern literature. Recognizing Wicksell effects is important to correct over-simple neoclassical parables, yes; but this is not to agree with the frequently met notion that, in consequence, the steady-state interest rate of perfect competition can lack *intertemporal Pareto-optimality* when Wicksell effects are present. Actually, no matter what 'reswitchings' or Wicksell effects are present, the competitive equilibrium does support *intertemporal production-and-consumption 'efficiency'*.

Fruitful critiques have been made in our time by Joan Robinson (1956) and Pierro Sraffa (1960) of the simple parable that lower steady-state interest rate *must* be associated with 'moreroundabout' modes of production. What remains intact is only this: (1) if a stationary population is endowed with capital goods that cannot support a golden-rule state of maximal per capital consumption, it can evolve into that golden-rule state *only by transiently sacrificing some current consumption in return for permanently enhanced consumption*; (2) for each specified interest rate, there is a convex tradeoff frontier between steady-state real factor prices (real wage, real rent of land, etc.), and any increase in that interest rate must shift downward that tradeoff frontier.

This first 1893 work illustrates Wicksell's virtues: his generosity in recognizing contributions of others; his confession that specified problems remain unsolved, or fail to be solved to the satisfaction of his scientific conscience; his depth of insight into the essence of an economic situation. Four decades before the mathematician Abraham Wald used inequalities and zero prices to ensure existence of a competitive equilibrium, Wicksell (1893a, p. 84, n.1) adumbrates the duality equalities-inequalities that common sense of economics requires. Wicksell (1934, pp. 180–81) gives pictures and words of a 'switch point', where two different activities that are coexistable are combined in any weighting just as ice and water coexist in any proportions at the freezing point. In his seventy-third year, Wicksell (1954) works out how fixed capital can be added to the Austrian models of circulating capital, an exposition that could be improved on only if he had replaced straight-line depreciation by the more convenient exponential depreciation.

Wicksell was an important creator of the neoclassical theory of the distribution of income according to the principles of marginal productivity. His work postdated J.B. Clark's breakthrough of the late 1880s; Phillip Wicksteed's conscious articulation in the early 1890s of first-degree-homogeneous production functions whose marginal products do exactly 'exhaust' the output; Léon Walras's mid-1890s generalization to infinite-many substitutable techniques of production, from his first edition's single technique and his second edition's finite-number of activities, a generalization that Walras could achieve only with the prior help of Pareto and Enrico Barone. If 'marginalism' is the essence of neoclassicism – and it is surely one important component – then Wicksell forms a trio with Johann Heinrich von Thünen (1826, 1850) and J.B. Clark (1899) as an archetypal neoclassicist.

Moreover, Wicksell's turn-of-century marginal productivity utilized the macromodel methodology of his 1893 capital theory: a simple general-equilibrium for society is envisaged with a single good and its production function. This is not macroeconomics in the modern Keynesian sense involving general price levels and elements of effective demand (the sense in which, up until now in this article, the word macroeconomics has been used). Rather, it is macroeconomics in the secondary sense that the word connoted in the mid-1940s when it early appeared in the literature: the sense of a Clarkian aggregate produced by aggregate labour and one idealized total of homogeneous capital. Indeed, the so-called 1927 Cobb–Douglas production function, $Q = L^k C^{1-k}$, was already buried in Wicksell's earlier writings of the Victorian era. (See also Wicksell 1934, pp. 125, 286.)

David Ricardo shocked his followers and contemporaries with a new chapter in the last edition of his *Principles*, which asserted that invention of machinery could harm wages and cause total production of society to shrink. A Wicksellian exposition, in which $Q(L, C)$ denotes output and $\partial Q(L,C)/\partial L$ denotes the real wage, clearly exposes the possibility that a viable invention which raises Q for fixed L and C can most certainly lower absolute $\partial Q(L,C)/\partial L$.

Uncharacteristically, Wicksell (1934, p. 137) blundered in falsely accusing Ricardo of error: under Ricardo's classical hypothesis that labour supply adjusts to keep the real wage near a constant level of subsistence, total Q could indeed be induced to shrink by a technology parameter's shift that lowered $\partial Q(L,C)/\partial L$ – as when raising the technical parameter T some what above unity near $(L,C,T) = (1,1,1)$ definitely does depress Q in $Q = 1.5\ (T-1)C + T^{-1}\ (LC)^{1/2}$ when the real wage is kept constant by downward adjustment of the labour supply.

Wicksell has good company in making this error: such eminent modern economists as Nicholas Kaldor and George Stigler, rightly impressed with the Pareto-optimality of competition's Invisible Hand in selecting viable inventions, wrongly infer that Ricardo's asserted drop in Q would contradict this Pareto Optimality and hence wrongly judge that there has to be an error on Ricardo's part. Since Ricardo is envisaging an induced drop in L, he is correct to assert that Q may well be decreased by the invention.

Of all the neoclassicists Wicksell is the most humanitarian, the least conservative. During his sixth decade of life he went to jail for the crime of

blasphemy. A friend of the *avant garde* August Strindberg, Knut Wicksell espoused redistribution from rich to poor at whatever cost of his own career. No writer of the Edwardian age came closer to the New Deal ideology of 1933–65 and to that of modern social democracy than did Wicksell. Yet, using the words '. . . the Hegelian darkness – and conceit of Karl Marx . . .', Wicksell (1893b, Preface, p. i) explicitly rejected Marxism as a paradigm to diagnose and understand the laws of motion of capitalism and as an erroneous programme for improving the welfare of the worker and peasant classes. His rejection of Marxism was based on knowledge of Marx's analysis and not on *a priori* prejudice; actually honesty in this regard inflicted a cost in terms of Wicksell's popularity, since Scandinavia was no exception to the rule that Marxism generated much political appeal in the three decades before World War I, the epoch just after Karl Marx's own death.

Despite his admiration for Ricardo, Wicksell denounced the unrealism of that writer's labour theory of value. Even the great editor of Ricardo, in Sraffa (1951, p. xxiii), lets that writer get away with transparent murder in fallaciously claiming to be able to 'get rid of [the complication of land and] rent' by setting each good's price to its labour cost on extensivemargin zero-rent land. Wicksell (1934, p. 24) points out that where the extensive margin for land will fall is itself an *endogenous* variable that is changed when the composition of demand alters away from land-intensive corn and toward labour-intensive cloth. Along with dozens of other self-contradictions in Ricardo's writings, there is clear recognition in his new chapter on machinery that a wartime shift of demand toward labour services of soldiers rather than toward rural produce would alter the distribution of income – a passage which is the root source for J.S. Mill's later overblown doctrine that 'demand for goods is not demand for labour'. If taste changes can alter distributive shares, then hopeless is Ricardo's attempt to separate distribution theory from value theory – and Wicksell was not loathe to call a spade a spade and a hopeless task hopeless.

One is left, most of all, with an impression of Wicksell's depth and breadth. Except for the Åkermans and Lindahl, Wicksell in his brief end-of-life professorship at Lund had almost no career-economist pupils. But it is no accident that Eli Heckscher and Bertil Ohlin should have originated the paradigm of factor–price equalization by free trade in goods. For, in his postretirement years back at Stockholm, Knut Wicksell was a national treasure who inspired a generation of younger economists (and succeeded, partially, in keeping Gustav Cassel scientifically honest).

One of the many harvests of his versatility is the 'voluntary exchange' (or 'benefit') theory of public finance and taxation. Wicksell [1986] began this Wicksell–Lindahl–Musgrave–Samuelson–Vickrey theory of pure public goods and the work of his pupil, Erik Lindahl (1919), created its foundation. When private goods consumed by a single person only are supplemented by a public good that is simultaneously enjoyed by many people, Pareto optimality required that production of the public good be carried to a point where its marginal (opportunity) cost just equals the sum of all citizen's marginal-rates-of substitution between the public good and any private good. Relying on a hoped-for Scandinavian consensus or 'unanimity', Wicksell perhaps worried too little about the 'free rider' problem (that results from the fact that every citizen in a Lindahl market is tempted to pretend not to much want the public good).

Finally, Wicksell's civility towards his great rival Gustav Cassel sets us all a noble example. Cassel had every gift except the gift of 'maybe'. Tutor to the King, Cassel pleased the Establishment and, prior to the post-1930 age of Keynes, was the economist most quoted by the international press. Although Schumpeter called Cassel '90% Walras and 10% water', I judge him to be a creative scientist, underrated today because of his egotistical failures to acknowledge doctrinal borrowings. (When his secretary wrote his biography, she alibied for this failing on the grounds that friction with his father blotted out from Cassel's memory all writers from whom he learned in early life!)

Wicksell (1934, pp. 219–52) gives his final reckoning with Cassel. His verdicts are unsparing, often harsh, but by no means malicious.

Moreover, by modern standards, often it is Wicksell who must be judged to have the weaker case. Cassel (1918) brilliantly anticipated the Harrod–Domar mode of balanced exponential growth, and deserves praise not blame for softpedalling diminishing returns to land in modern Europe. Cassel, like Pareto, was right to downplay cardinal measurable marginal utility even though he went too far in hypothesizing reduced-form, positivistic demand functions. Wicksell was not wrong in wishing for individual tastes to underlie welfare economics; but, as Abram Bergson (1938) was later to demonstrate, ordinal welfare economics does not necessarily require independently addable Benthamite utility functions for individuals. Wicksell, better than Marshall or Mises or Walras, realized from the beginning that competitive equilibrium of *laissez faire* does not necessarily achieve or approximate to a state of maximal social welfare or equity. He recognized, dimly, that the algorithm of perfect competition (no to be confused with *laissez faire*) does achieve the efficiency of production and exchange that we have since 1950 called 'Pareto optimality'; so, with the aid of feasible-best prior redistribution of people's endowments, the competitive market mechanism might be used to contrive a state of ethical optimality. Better to see obscurely in 1893 what we came to understand only after 1938. But, short of that, better for Wicksell to stubbornly insist that the emperor of market competition wore no ethical cloths than fall in with capitalistic apologias.

Wicksell's economics, because of its eclecticism and generality, adapts well to the present post-neoclassical age. As with Cournot, his writings speak eloquently to readers of a later century.

Bibliography

Bergson, A. 1938. A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52: 310–334.

Cassel, G. 1918. *Theory of social economy*. Trans., New York: Harcourt, Brace & Co., 1932.

Clark, J.B. 1899. *The distribution of wealth*. New York: Macmillan.

Gårdlund, T. 1956. *The life of Knut Wicksell*. Trans. N. Adler. Stockholm: Almqvist & Wiksell, 1958.

Hicks, J.R. 1932. *The theory of wages*. London: Macmillan.

Patinkin, D. 1982. *Anticipations of the general theory? And other essay on Keynes*. Chicago: University of Chicago Press.

Radcliffe, Lord (Chairman). 1959–60. *Committee on the working of the economy: Report, minutes of evidence, and memoranda*. London: HMSO.

Robinson, J. 1956. *The accumulation of capital*. London: Macmillan.

Samuelson, P.A. 1983. Thünen at 200. *Journal of Economic Literature* 21: 1468–1488.

Stigler, G.J. 1941. *Production and distribution theories: The formative period*. New York: Macmillan.

Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.

Sraffa, P. (With the collaboration of M.H. Dobb.) 1951. General Preface to D. Ricardo, *On the principles of political economy and taxation*. Cambridge: Cambridge University Press for the Royal Economic Society.

von Thünen, J.H. 1826, 1850. *Der Isolierte Staat*, Pts I and II.

Wicksell, K. 1893a. *Über Wert, Kapital und Rente*. Jena: Gustav Fischer. Trans. as *Value, capital and rent*. London: George Allen & Unwin, 1954. This contains a fairly complete Wicksell bibliography, on pp. 169–175, completed by Arne Amundsen.

Wicksell, K. 1893a. *Finanztheoretische Untersuchungen nebst Darstellung und Kritik des Steuerwesens Schwedens*. Jena: Gustav Fisher.

Wicksell, K. 1898. *Interest and prices*. Trans., London: Macmillan, 1936.

Wicksell, K. 1934. *Lectures on political economy*. Vol. I: *General theory*; Vol. II: *Money*. London: George Allen & Unwin.

Wicksell Effects

Edwin Burmeister

Keywords

Capital deepening; Capital goods; Capital theory; Capital theory paradoxes; Numeraire; Production possibility frontier; Wicksell effects; Wicksell, J. G. K.

JEL Classifications

D5

In realistic economic models with n different types of capital goods, the value of the capital stock is

$$V = \sum_{i=1}^{n} P_i K_i \tag{1}$$

where P_i is the price of the ith capital good in terms of some *numéraire*. The value of capital, however, is not an appropriate measure of the 'aggregate capital stock' as a factor of production except under extremely restrictive conditions. Wicksell (1893, 1934) originally recognized this fact, which subsequently was emphasized by Robinson (1956).

If attention is restricted to alternative steady-state comparisons, in constant-returns- to-scale economies without joint production V is a function of the interest rate, r; see, for example, Burmeister and Dobell (1970). The *Wicksell effect* is the change in the value of the capital stock from one steady state to another, namely

$$\frac{\mathrm{dV}}{\mathrm{dr}}. \tag{2}$$

The term 'Wicksell effect' was introduced by Uhr (1951), but its importance was not widely recognized until the writings of Robinson (1956) and Swan (1956).

The Wicksell effect is the sum of the *price Wicksell effect* (which is the revaluation of the inventory of capital goods due to new prices) and the *real Wicksell* effect (which is the price-weighted sum of the changes in the physical quantities of different capital goods):

$$\frac{\mathrm{dV}}{\mathrm{dr}} = \sum_{i=1}^{n} \frac{\mathrm{dP}_i}{\mathrm{dr}} K_i + \sum_{i=1}^{n} P_i \frac{\mathrm{dK}_i}{\mathrm{dr}}. \tag{3}$$

Numerical examples show that the price Wicksell effect can be negative, that is,

$$\sum_{i=1}^{n} P_i \frac{\mathrm{dK}_i}{\mathrm{dr}} < 0 \tag{4}$$

is possible, even when (i) the total Wicksell effect is positive [$\mathrm{d}V/\mathrm{d}r > 0$] , or (ii) particular capital stocks are increasing with $\mathrm{d}K_i/\mathrm{d}r > 0$ for some but not all i; see Burmeister and Dobell (1970, pp. 289–93). In neoclassical models with only one capital good ($n = 1$), the real Wicksell effect is always negative. Moreover, the sign of the price Wicksell effect depends upon the choice of *numéraire*, and hence so does the total Wicksell effect given by (3). The sign of the real Wicksell effect, however, is independent of the choice of *numéraire*.

One central issue of the Cambridge controversies in capital theory involves Wicksell effects. Does a decrease (increase) in the steady-state interest rate always imply a rise (fall) in per capita steady-state consumption provided the rate of interest is greater (less) than the rate of growth of labour? In one-capital good models, the answer to this question is, 'Yes'. In general, the answer is, 'Yes', if and only if the real Wicksell effect is negative; see Burmeister and Turnovsky (1972) and Burmeister (1976).

To establish this relationship between the behaviour of per capita consumption and the real Wicksell effect, consider a technology which can be represented by a *production possibility frontier*

$$Y_1 = T(Y_2, \ldots, Y_n; L, K_1, \ldots, K_n) \tag{5}$$

where Y_i is the output of commodity i, L is the labour which grows at the exogenous rate g, and K_i is the stock of commodity i used as a capital input.

It is assumed further that $T(\cdot)$ is twice continuously differentiable, exhibits constant returns to scale, and has a Hessian matrix $[T_{ij}]$ that is negative semi-definite and whose rank varies with the degree of joint production in the economy; see Samuelson (1966), Burmeister and Turnovsky (1971), and Kuga (1973). The analysis which follows can be generalized to non-differentiable technologies as in Burmeister (1976), but for simplicity only differentiable technologies are considered here.

In steady-state equilibria all quantities grow at the rate g, implying that the output of every commodity must satisfy

$$Y_i \equiv C_i + K_i = C_i + gK_i, \quad i = 1, \ldots, n, \tag{6}$$

where C_i denotes the consumption of commodity i. Substituting these steady-state restrictions into (5) and using lower-case letters to denote per capita quantities, we have

$$c_1 + gk_1 = T(c_2 + gk_2, \ldots, c_n + gk_n; 1, k_1, \ldots, k_n). \tag{7}$$

Let the prices of commodities and the rental rates for capital goods, both in terms of the wage rate as *numéraire*, be denoted by p_i and w_i respectively, $i = 1, \ldots, n$; also let r denote the interest or profit rate. It is well-known that intertemporal profit maximization and/or efficiency necessitates that

$$\frac{\dot{p}_i}{p_i} + \frac{w_i}{p_i} = r, \quad i = 1, \ldots, n. \tag{8}$$

Imposing the steady-state requirement that relative prices remain constant, (8), implies that

$$w_i = rp_i, \quad i = 1, \ldots, n. \tag{9}$$

Using the well-known marginal conditions

$$\frac{\partial T}{\partial(c_i + gk_i)} = -\frac{p_i}{p_1} \text{ and } \frac{\partial T}{\partial k_i} = \frac{w_i}{p_1}, \quad i = 1, \ldots, n, \tag{10}$$

we see that a vector

$$(c^*, k^*; r^*, p^*) = (c_1^*, \ldots, c_n^*; r^*, p_1{}^*, \ldots, p_n{}^*) \geq 0 \tag{11}$$

satisfying (7) and (10) represents a steady-state solution at the growth rate g. It thus follows immediately from differentiation of (7) that almost everywhere.

$$\sum_{i=1}^{n} p_i \left(\frac{\mathrm{dc}_i}{\mathrm{dr}}\right)|_{(r^*, p^*)} = (r - g)\sum_{i=1}^{n} p_i \left(\frac{\mathrm{dk}_i}{\mathrm{dr}}\right) \times|_{(r^*, p^*)}; \tag{12}$$

see Burmeister (1976) for details.

Now let v denote the per capita value of capital in terms of the wage rate as *numéraire*:

$$v = \sum_{i=1}^{n} p_i k_i. \tag{13}$$

The change in the per capita value of capital across alternative steady-state equilibria is obtained by differentiating (13); thus almost everywhere the per capita Wicksell effect is

$$\frac{\mathrm{dv}}{\mathrm{dr}}|_{(r^*, p^*)} = \sum_{i=1}^{n} \left(\frac{\mathrm{dp}_i}{\mathrm{dr}}\right)|_{(r^*, p^*)} . k_i + \sum_{i=1}^{n} p_i . \left(\frac{\mathrm{dk}_i}{\mathrm{dr}}\right)|_{(r^*, p^*)} \tag{14}$$

Comparing (14) and (12), it is seen that it is the real Wicksell effect which determines whether or not 'consumption' is well-behaved across steady-state equilibria. That is, if the real Wicksell effect is negative and

$$\sum_{i=1}^{n} p_i . \left(\frac{\mathrm{dk}_i}{\mathrm{dr}}\right)|_{(r^*, p^*)} < 0, \tag{15}$$

then almost everywhere

$$\sum_{i=1}^{n} p_i . \left(\frac{\mathrm{dc}_i}{\mathrm{dr}}\right)|_{(r^* p^*)} \gtreqless 0 \text{ as } r \gtreqless g. \tag{16}$$

In particular, when $c_2 = c_3 = \ldots = c_n$ and only commodity 1 is consumed, consumption as measured by c_1 rises (falls) as r is increased from r^* to $r^* + \Delta r^*$ when $r^*$is greater (less) than g. (The familiar golden rule condition giving maximum per capita steady-state consumption holds at $r^* = g$).

It follows, then, that a negative real Wicksell effect is the appropriate concept of 'capital deepening' in a model with many heterogeneous capital goods. That is, when (15) and hence (16) hold, an economy with a low interest rate (but exceeding g) has 'more capital' than one with a higher interest rate in the sense that it is capable of providing more steady-state per capita

consumption. Although (15) and (16) always hold in a neighbourhood of $r^* = g$, examples show that they do not generally hold everywhere. This possibility – that (16) does not hold everywhere – is perhaps the most interesting conclusion to emerge from the Cambridge controversies and has been termed a 'paradox'. However, the 'paradox' involves comparisons of alternative steady-states rather than comparisons of alternative feasible paths; Bliss (1975) provides a lucid explanation of why such 'paradoxes' are in fact not surprising or damaging to the neoclassical paradigm.

Imposing some set of conditions on the technology $T(\cdot)$ should be sufficient to assure that the real Wicksell effect is always negative. Such conditions would be of interest – especially if they could be empirically tested – since they would validate the qualitative conclusions derived from the one-good models often used in macroeconomics without any theoretical justification for ignoring capital aggregation problems. Moreover, Burmeister (1977, 1979) has proved that a negative real Wicksell effect is a necessary and sufficient condition for the existence of an index of capital, κ, and a neoclassical aggregate production function $F(\kappa)$ defined across steady-state equilibria such that (i) $c = F(\kappa)$, (ii) $r = F'(\kappa)$, and (iii) $F''(\kappa) < 0$. Unfortunately, no set of such sufficient conditions is known, but the literature on capital aggregation suggests that they would impose severe restrictions on the technology.

See Also

- ▶ Capital Theory (Paradoxes)
- ▶ Reswitching of Technique
- ▶ Reverse Capital Deepening

Bibliography

Bliss, C. 1975. *Capital theory and the distribution of income*. Amsterdam: North-Holland.

Burmeister, E. 1976. Real Wicksell effects and regular economies. In *Essays in modern capital theory*, ed. M. Brown, K. Sato, and P. Zarembka. Amsterdam: North-Holland.

Burmeister, E. 1977. On the social significance of the reswitching controversy. *Revue d'économie politique* 87: 330–350.

Burmeister, E. 1979. Professor Pasinetti's 'unobtrusive postulate', regular economies, and the existence of a well-behaved aggregate production function. *Revue d'économie politique* 89: 644–652.

Burmeister, E., and A.R. Dobell. 1970. *Mathematical theories of economic growth*. New York: Macmillan.

Burmeister, E., and S.J. Turnovsky. 1971. The degree of joint production. *International Economic Review* 12: 99–105.

Burmeister, E., and S.J. Turnovsky. 1972. Capital deepening response in an economy with heterogeneous capital goods. *American Economic Review* 62: 842–853.

Kuga, K. 1973. More about joint production. *International Economic Review* 14: 196–210.

Robinson, J. 1956. *The accumulation of capital*. London: Macmillan.

Samuelson, P.A. 1966. The fundamental singularity theorem for non-joint production. *International Economic Review* 7: 34–41.

Swan, T. 1956. Economic growth and capital accumulation. *Economic Record* 32: 334–361.

Uhr, C.G. 1951. Kunt Wicksell, a centennial evaluation. *American Economic Review* 41: 829–860.

Wicksell, K. 1893. *Value, capital and rent*. Reprinted New York: Augustus M. Kelley, 1970.

Wicksell, K. 1934. *Lectures on political economy*. 3rd ed. London: George Routledge and Sons Limited, 1938.

Wicksell, Johan Gustav Knut (1851–1926)

Carl G. Uhr

Keywords

Aggregate demand and supply; Böhm-Bawerk, E. von; Capital theory; Cassel, G.; Davidson, D.; Malthus's theory of population; Marginal cost pricing; Marginal productivity theory of distribution; Marginal utility theory; Market socialism; Monetary equilibrium; Period of production; Socialism; Stockholm School; Wicksell effects; Wicksell, J. G. K.

JEL Classifications

B31

Life and Career

Johan Gustav Knut Wicksell was born in Stockholm on 20 December 1851, the youngest of six children of Johan and Christina Wicksell. One child died in infancy, so Knut grew up with three sisters a few years older than he, and a brother, Axel, one year older.

Knut's mother died when he was not quite seven, an event that greatly affected the sensitive boy. His father, a produce dealer, remarried in 1861 but died five years later when Knut was 15. After that the children moved to live for a time with an aunt and their maternal grandmother. In the last decades of his life Knut's father had become moderately well-to-do by investing profits from his grocery business in rental properties. The estate that was left at his death yielded an income sufficient to provide for the children and their education through gymnasium (high school), and for the two boys a start at the University of Uppsala.

At the gymnasium Knut had already shown considerable aptitude for languages and an unusual ability at mathematics. Thus, when he enrolled at Uppsala University 1869, it was with the intention of becoming a mathematician with physics as a second field of study.

From about age 15 Knut came increasingly under the influence of a pietistic pastor of the Swedish Lutheran Church. This religious phase lasted about seven years, in the course of which he became a devout Christian; he withdrew from most social activities to study the Bible and meditate. At the same time he made rapid progress in his studies of mathematics, physics and astronomy, earning his first degree, BS *cum laude,* in 1871, after only two rather than the usual four years at the university, and then proceeding to graduate studies.

However, doubts had begun to assail his faith, and in spring 1874 he had an emotional crisis from which he emerged, and for the rest of his life remained, a free thinker. He became a strictly a-religious philosophical rationalist who, later on, became known as an outspoken and witty critic of the Christian religion in all its forms.

Until 1873 Wicksell maintained himself at the university on the modest annual income he received as his share of his father's estate, on a small inheritance from his grandmother, and on a succession of grants from private foundations. Now the last were drying up and the money from his grandmother was nearly gone. To add to his scant resources, he filled a vacancy as a teacher at a secondary school at Uppsala, 1873–4. The year after that he worked as a private tutor to the son of an ironmaster. Also from time to time he borrowed various amounts from one of his sisters who had established herself in Stockholm as a physiotherapist.

In fact, Wicksell's financial condition remained precarious and often severely strained, except for the years 1885 and 1888–9 when he was studying abroad largely supported by grants, for most of his adult life until 1901. Then, at age 50 and supporting his wife and two school-age sons, he was finally appointed professor extraordinarius (about equivalent to associate professor) at Lund University, and then, from 1904, he served there as ordinarius or full professor for 12 years, until his retirement in 1916.

In 1875 he passed two of three required examinations for the degree *philosophiae licentiatus* in mathematics (the *phil. lic.* is a graduate degree taken prior to the student's beginning work on the doctoral dissertation). Soon after that he began to doubt that he would be able to make any significant contributions to mathematics. While contemplating a change of career either to humanities or to the emerging social sciences, he immersed himself, over a long transition period, in the activities of the students' organization, the Student Corps. He was elected as its curator, 1877–9. In that post he became well known for his critical social views and for his surprising effectiveness as a speaker. At this time he also wrote some 'social indignation' poetry as well as some plays, one of which proved popular and was performed at Uppsala and also in some other towns.

In 1879 two events, in themselves inconspicuous, occurred which strongly influenced Wicksell's subsequent career. He moved to share an apartment with two advanced graduate students, H. Ohrvall in medicine, and T. Frölander in law, and he acquired a book just recently

released in Swedish translation, G. Drysdale's tome, *The Elements of Social Science,* with its challenging subtitle, 'Physical, Sexual, and Natural Religion; An Explanation of the True Causes and Cure of the Three Primary Evils of Society -Poverty, Prostitution, and Celibacy'. This work, published in England 1854, became very popular in the Swedish translation of 1878, and went through over 30 reprintings over the years.

The three men, whose outlook on society was in several respects similar, became lifelong friends. What cemented their friendship was that they set about on their own and jointly to study Drysdale's thoroughly neo-Malthusian treatise. It discussed frankly several subjects then regarded as unmentionable in 'polite society', such as sex, methods of birth control, the allegedly harmful psychological effects of celibacy if continued for a decade or more past puberty, prostitution as the only alternative for the young among the poor, the need for family planning, and the need to limit population growth in order to raise the standard of living for the working class above bare subsistence.

Wicksell treated this book as a revelation. It focused his mind on 'the social question', that is, on the social sciences towards which his inclination guided him more and more. As an early result of studying Drysdale, supplemented by some writings of J.S. Mill, in February 1880 he gave a lecture to a temperance lodge at Uppsala on 'The Most Common Causes of Habitual Drunkenness and How to Remove Them'. His address got a mixed reception but was reported in the local newspaper, which led to an insistent demand for him to repeat his lecture two weeks later in a much larger hall which was filled to overflowing.

There he attributed alcoholism, widespread among factory workers, to the poverty and monotony of their lives, with wife and children crammed into crowded and often insanitary housing. For this the local inn offered almost the only relief and relaxation available. And with the young workers this led to the use of the services of prostitutes since these workers for years earned too little to marry and start a family. The remedies he urged were for the medical profession to be assigned the duty of disseminating information about birth control techniques, and for the public health authorities to set and enforce standards of sanitation and room space per occupant in housing in the factory districts of cities and towns. The reaction to his lecture was strong. Papers by the Young Socialists and by some student organizations praised him. Medical and temperance organizations either reviled or ridiculed him, and several newspapers questioned his competence to pronounce on some of the sensitive issues he had covered.

From now on the die was cast. There would always be one or more reporters present at his future appearances, because these were certain to be newsworthy. Reporters would summarize his talks and write longer accounts about how his audience reacted, especially the critics and opponents among them, and how he, in turn, responded to critics. Most of the reportage depicted him as a non-revolutionary radical social reformer, and that was how public opinion came to view him. We may add that he himself did nothing to modify and much to strengthen that impression.

Later in 1880 he issued his lecture as a tract of some 90 pages and along with it a pamphlet, 'Answers To My Critics', both of which sold in several thousand copies. In fact, this became something of a pattern. Between 1880 and 1885, and again in 1886–7, after his return from his first stay abroad, Wicksell had in substance turned into a radical public lecturer and journalist. This was how he earned his spartan maintenance, by paid public lectures sometimes followed by publishing tracts based on them, and by paid articles written in neo-Malthusian spirit on various 'social questions' for several, sometimes in a given week for as many as ten, different city and town newspapers.

In 1885 he set aside his journalistic work for a time and completed the last requirement for the *phil. lie.* degree in mathematics by a research paper, the other requirements he had met in 1875. Now, however, he wanted to shift into the social sciences rather than go on for a doctorate in mathematics. To do that at any level higher than the elementary meant study abroad, for at that time the social science disciplines were not separate fields but were elements of the curricula in

law, philosophy, the humanities or theology in Sweden's universities. But he had no funds for going abroad. Then help came unexpectedly.

His sisters had an opportunity to sell the rental properties of the Johan Wicksell estate to a buyer at a favourable price if Knut and his brother, Axel, who had emigrated to the United States, would agree, as they did. Knut's share of the proceeds was sufficient to pay off his old debts and also to maintain him for about a year abroad, and so in autumn 1885 he went to London.

In London Wicksell spent his days studying some of the classical economists and treatises by Cairnes, Jevons, Walras and Sidgwick, his first exposure in depth to economics, and his weekends in meetings with persons to whom he was introduced by Charles Drysdale, an engineer who continued the neo-Malthusian activities his father, George Drysdale, had initiated. Thus he met prominent British neo-Malthusians, Annie Besant among them, Karl Kautsky and some labour leaders, and the leading Fabians.

By summer 1886 he returned to Uppsala and Stockholm to resume his public lecturing, writing for several newspapers, and composing tracts. This was now a matter of necessity, for he had used up his patrimony during his stay in Britain. In 1886–7 he delivered 42 public lectures in towns in central Sweden, in Copenhagen and Christiania, for fees which paid very little above his travel and maintenance expense. The subjects he spoke on were as follows:

Marriage	14 lectures
Population control	10 lectures
Socialism	6 lectures
Prostitution	5 lectures
Spiritualism	2lectures
Why not a free-thinker?	2 lectures
Religion	1 lecture
Euthanasia	1 lecture
Impression of Britain	1 lecture

At the end of 1885 Victor Lorén, a wealthy young man, greatly interested in promoting the social sciences after studying them in Germany with Roscher, bequeathed his estate to a foundation bearing his name, with instructions that it should be used for the promotion of studies and research and publications by scholars devoted to economics and related social sciences. Wicksell was still in London when early in 1886 he was informed that the Loren Foundation was awarding him a grant for up to three years to study economics at universities in Germany and Austria. Loren's relatives unsuccessfully contested his will in court, but this held up the grant until the summer of 1887, when the suit was settled.

If the Loren Foundation had not given him that large grant (and later smaller ones for each of the five treatises he published between 1893 and 1906) Wicksell could hardly have become an economist, much less a major figure in this discipline. As it was, he went first to London to renew acquaintances. In October 1887 he went to the University of Strassburg to follow lectures by Brentano on labour economics, on money and credit by both Brentano and Knapp, and on economic distribution by Singer.

In spring 1888 he was in Vienna to listen to Carl Menger's lectures. In July he returned for a short stay in Sweden. On his way there he met Anna Bugge, a Norwegian gymnasium teacher, who later became his wife. By autumn 1888 he was at the University of Berlin to follow the lectures of Adolph Wagner on public finance. In spring 1889 he returned to Sweden to seek a lectureship in economics at the University of Stockholm. He was turned down as being 'too notorious' a person. Summer 1889 he decided to spend the rest of his grant studying economics in Paris. Before going there he took a trip to Christiania to see Anna Bugge, with whom he had corresponded while in Germany. There he proposed a common-law marriage to her, but out of consideration for her parents she turned him down, whereupon he left in a huff for Paris.

A word may be needed here about the romantic side of Wicksell's life. It is known that he was infatuated in his early twenties with two young ladies. But he was always shy and very hesitant in socializing with young women. So the first young woman moved to Switzerland and married there. The second one was a case of love at a distance, for he failed even to make contact with her. The third and last incident occurred years later. For a part of the summer 1886 he was invited to stay in

his friend Frölander's household in Stockholm, where he gave most of his lectures. But there he soon found himself becoming infatuated with his friend's wife. So before he might say or do something to jeopardize their friendship, he made the proper excuses and returned to his lonely lodgings in Uppsala.

Anna, however, did not want to give up Knut. She decided she would be happier with than without him even at the cost of estrangement from her parents. She joined him in Paris that summer; he was then 37 and she 26.

In Paris he attended lectures on public finance by Leroy-Beaulieau and on population theory by Desmoulin, and he began to publish in economics. His first article, the translated title of which is 'Empty Stomachs–Full Stores' came out in a Norwegian journal *Samtiden* in 1890. His second article, 'Überproduktion oder Überbevölkerung' (Excess production or excess population) appeared in *Zeitschrift für die gesamten Staatswissenschaften,* also in 1890. In both he argued that it was fluctuations in capital formation that made the difference between prosperity and depression. In recovery a rate of capital formation is generated which fails to be sustained because consumption demand, though rising, lags behind the rate at which productive capacity expands on a growing capital base.

In summer he and Anna returned to Stockholm. Though soon to be a father (their first son, Sven, was born in October 1890, and a second son, Finn, in 1893), Wicksell had no settled way of earning a living. Economics was then taught only in the faculties of law. Those teaching it in addition to a doctorate in economics also had to have at least an undergraduate degree in law in order to give courses on law and economics as related mainly to taxation and public finance. So he had no alternative but to return to being a freelance journalist and public lecturer.

During the years 1890–9 Wicksell had more trials and tribulations, only a few of which can be related here. He gave rather few public lectures, but some had a very negative effect on his public image.

In 1892 the government wanted to increase the duration of the compulsory military service to strengthen the country's defences. In November Wicksell lectured in Stockholm on the question, 'Can Sweden Protect her Independence?' He argued, and most of his listeners might have agreed with him, that no matter how long the draft were extended, it would not be adequate for defending Sweden against attack by a major military power. But they disagreed vehemently when he went on to say that since the country could not defend itself on its own resources, it would make better sense to disarm and use the resources set free from defence for other domestic purposes. Then Sweden ought to negotiate for incorporation into the Russian empire with its much greater military resources. In return for the protection thus provided, the Swedes with their long traditions of democracy ought then to play a civilizing role within and for the Russian empire.

This performance earned him the sobriquet of 'defence nihilist', which did not deter him 12 years later when another draft extension was proposed from repeating this same lecture, May Day 1904. At that time it occasioned even greater offence and ridicule than in 1892.

His article 'Kapitalzins und Arbeitslohn' (Interest and wages), published in the *Jahrbücher für Nationalökonomie und Socialwissenschaft und Statistik* 1892, formed the basis for the marginal productivity theory of distribution – one of Wicksell's main contributions to economic theory – which he developed in his first treatise, *Uber Wert Kapital und Rente* 1893 *(Value, Capital and Rent,* translated 1934). This remarkable work received initially almost no attention in Sweden, but was favourably reviewed by both Böhm-Bawerk and Walras.

Next he turned to an examination of Sweden's taxes in his popular tract, *Our Taxes–Who Pays and Who Ought to Pay Them?* (99 pp., 1894), issued under the pseudonym of Sven Trygg. He was outraged at the regressiveness of the country's taxes. That, he concluded, had to be a consequence of the fact that only the well-to-do could vote, as income and property qualifications for the franchise excluded almost all the workers and most of the small farmers.

The analysis of that tract was extended and refined in his second treatise, *Finanztheoretische*

Untersuchungen (Studies in the theory of public finance), 1896. There he urged that the major part of the revenue burden be shifted from indirect to direct progressive taxes on income and wealth. That treatise also embodied his design on an 'equitable' tax system based on an application of marginal utility theory to the public sector, and a methodology (essentially marginal cost pricing) for pricing pure and less than pure public goods, the services of public utilities, and the products of market-sharing oligopolies and cartels.

In fall 1894, Wicksell applied at Uppsala University to have *Value, Capital and Rent* evaluated as a doctoral dissertation. The answer was 'no', with the added advice to use it for a viva voce examination of a *phil. lic*. degree in economics. David Davidson was appointed examiner and Wicksell passed with high marks in May 1895. Next he needed the doctorate. In 1896 he submitted the first part of his *Finanztheoretische Untersuchungen,* 'Theory of incidence of taxation', as a dissertation. Again Davidson was chief examiner, and the degree was awarded Wicksell *magna cum laude.*

That done, he began research on monetary theory and policy, which he completed as his third treatise, *Geldzins und Güterpreise,* 1898 *(Interest and Prices,* translated 1936), the home of the Wicksellian 'cumulative price level fluctuations or processes', allegedly generated by a divergence between the rate of return on newly created real capital and the bank-dominated market rate of interest.

Now he applied both at Stockholm and Uppsala universities for a docentship but was rebuffed because he lacked a degree in law. From 1890 into 1897 he had maintained his family slightly above subsistence level by earnings from his newspaper articles and tracts and from a succession of Lorén grants. However, in autumn 1897 he decided at real hardship, with no more Lorén money, to move from Stockholm to Uppsala to devote his entire energy to cramming through law courses as fast as possible to a *juris candidatus* of LL. B. degree. To do this he had to maintain his family by borrowing from his friends Ohrvall, a physician, and Frölander, a banker-lawyer, both of whom were doing well. In 1899, in less than two years, he had earned his law degree, which usually takes undergraduates four years. He was appointed a docent at Uppsala University but without fixed salary. Consequently his income depended on how many law students came at a given fee per head to attend his tutorials.

At the Lund University faculty of law a professorial vacancy was created when an older professor's post, viewed as overloaded, was split to shift its courses in tax law and economics from the old position to the new one. But Parliament, in approving this, had voted less money for it than a full professor's salary. Wicksell and three others, including Gustav Cassel, competed for this post of professor extraordinarius. As the other candidates (Cassel for lack of a law degree) were eliminated as not sufficiently qualified, the appointment was offered to Wicksell in January 1900. For complex reasons the upgrading of this to ordinary or full professorship was delayed until January 1904, when Wicksell, at the age of 53, was finally securely established as a full professor.

At Lund, where his teaching of tax law courses required much more preparation than economics, he still found time to write *Föreläningar i Nationalekonomi (Lectures on Political Economy I,* 1901, translated 1934). *Lectures I* were an expansion and improvement, especially in capital theory, over what he had presented in *Value, Capital and Rent.*

His courses in law as related to taxation were well attended but those in economics attracted very few students, it being an elective subject. He soon found out that the students lacked the background to get much out of a semester on *Value, Capital and Rent* and another on *Interest and Prices.* So he shifted his presentation from pure to applied economics to subjects such as agriculture and industry, commerce and consumption, social movements, social insurance, economic crises and inflation.

He had good relations with students. His approach to them was friendly. They, in turn, liked or were amused by his idiosyncrasies, and they admired his courage to fight for his convictions.

Unlike most professors, who at that time lectured in formal dress, swallowtail coats and all, Wicksell appeared in ordinary, rarely well-

pressed, street clothing. Instead of a top hat or a Derby he wore a visored cap, much like a fisherman's. Since he lived some distance from the university, he did the family's marketing at the nearby open-air market before his morning lectures. Consequently, as he strode in to the lecture room, he would adorn one side of the lectern or the other with his market basket filled with produce, meats and fruits.

In 1905 he issued one of his best and last tracts, *Socialiststaten och nutidssamhället* (The socialist state and contemporary society, 40 pp.). He restated more systematically his perspective on socialism which he had lectured on in the 1880s. First he made it clear he considered a limited but not a complete achievement of a socialist economy (with all means of production other than labour collectively owned and administered) to be inevitable in the future. Under universal adult suffrage the workers would be the political majority. As such, they would not for long tolerate the great inequalities of income and wealth and the economic instability (of employment and economic insecurity and dependence in old age) of laissez-faire capitalism without seeking and taking remedial measures.

He warned against drastic measures of income redistribution taken by a workers' government suddenly come to power. That would only yield a temporary gain followed by loss as private capital accumulation would all but cease before the workers' regime would have developed the means to replace it by public accumulation. A socialist economy is best built gradually by peaceful means and under democratic governance. Nationalization initially of natural monopolies and cartels might suffice if followed by substantial expansion of tax supported social security and social insurance schemes. For the sake of efficiency, he held it was best to leave farming and most varieties of genuinely competitive enterprises in private and/or cooperative ownership.

Consequently he argued for a form of market socialism with a well developed welfare state. It is surprising to recognize the great extent to which his social vision has become a reality in Sweden (and in Scandinavia as a whole) after more than half a century of Social Democratic rule.

In 1906 Wicksell published *Lectures on Political Economy II,* the volume on money and credit. In part an expansion and revision of what he had put forth in *Interest and Prices,* yet *Lectures II* were much more than that. They were epochmaking less for their particular findings than for the broad framework and methodology they provided for analysis of money and credit. *Lectures II* were translated first into German in 1922 when, in the midst of the German hyperinflation they were read with greater than usual interest, and into English in 1935.

Wicksell's years at Lund were very productive. He wrote about 50 articles and took an active part in the tax reform of 1910, in the national pension legislation of 1913, and, after the outbreak of the First World War, along with Davidson, he played an important role in the legislation and policies relating to banking, currency and exchange controls.

His work had continued in a tranquil manner until 1908. Then a young 'anarchist agitator' was sentenced to prison for 'disturbing the religious peace' by public blasphemy. He had published a parody of the Wedding at Cana in a socialist newspaper. His case, and two or three similar ones that had occurred earlier, impressed Wicksell as infringements by the courts of freedom of speech and press, guaranteed by the Swedish constitution. Against better advice he decided to make a test case of himself. Accordingly in November 1908 at Stockholm he lectured to a large audience on 'The Throne, the Altar, the Sword, and the Bag of Money' in which, inter alia, he satirized the story of the Immaculate Conception. He raised and answered the question:

> Why was not Joseph, the betrothed of the Virgin Mary, rather than the Holy Ghost allowed to father Jesus? Because then the world could not have been saved! Joseph's rights as an individual had to be set aside for the salvation of the many millions of souls in past centuries who would otherwise have gone to perdition and the further millions now and for all time to come until the Last Judgement.

Wicksell was tried and sentenced, against the protests of Social Democrats, organized workers and liberals, to two months in prison. He was allowed to select the jail where he would serve

his time. Early in 1910, after a higher court had sustained the lower court's decision, he chose the jail, known to be better than most, at the small fisherman's town of Ystad in southern Sweden. There he suffered no hardship. His university salary was withheld as long as he was a guest of the government. He used his time to advantage by writing his last tract, *Laran om befolkningen, dess sammänsattning, och förändringar* (The Theory of Population, its Composition, and Models of Change, 1910, 52 pp.).

There, apart from the clear demographic analysis it presented, he reiterated the conclusion from his public lectures of the 1880s, that, because of partial depletion and increasing scarcity of natural resources (in Sweden's case primarily timber and iron ore), the country's optimum population should be three million instead of its five million inhabitants, and for Europe a reduction to three quarters of its population as of 1910. Like Malthus and many other writers on population, while he acknowledged the productivity increasing effect of technological progress, he failed to see, and greatly underestimated the fact, that some of the new technology virtually adds to existing resources, in part by turning former waste products to productive uses, in part by increasing the number of uses to which existing resources can be turned.

Wicksell's remaining years at Lund passed quietly. But as his retirement was approaching it threatened renewed hardships for him and Anna. Before coming to Lund they had no savings, and when leaving, they had very little more than their household effects in rented housing. Since Wicksell had served only 16 years at the university, compared with colleagues who at age 65 had usually served 25 or more years, he was barely entitled to two-thirds of the usual professional pension. As the First World War inflated prices, especially in Stockholm to which city he and Anna insisted on moving, two-thirds pension would not pay for much more than house rent.

Two of his friends who were members of parliament succeeded on a motion to obtain a supplementary allowance for him which raised his pension to 90 per cent of the usual amount. There still remained the problem of housing, which had become very expensive in the capital. So his two parliamentarian and several other close friends, including David Davidson and Eli Heckscher, gathered together and by their personal contributions they raised enough money to buy a lot for a house and garden in Morby, a suburb of Stockholm, and to initiate construction to Anna's specifications. To complete the building of the house, Wicksell negotiated a small mortgage. By Christmas 1916 he and Anna moved into the first house they could call their own.

Now, in his last decade, a new phase of life began for both of them. Anna, who had taken a law degree in 1911 at Lund and had become a leader in the suffrage movement and later the peace movements, now had greater opportunity to be effective than at Lund. Wicksell, freed both from financial worries and the teaching of law courses, could devote himself full time to research and professional activities as an economist with the much greater resources at his disposal for research and opportunities for consultation of Stockholm as compared with Lund. The years 1917–26 were probably the most satisfactory and happiest in their lives.

Wicksell soon became very active. He wrote 29 articles from Morby on wartime inflation and how to roll it back, on Scandinavia's post-war monetary problems, and on capital theory. From 1915 he had been a consultant to the governor of the Bank of Sweden. In 1916 he and Davidson were appointed to a parliamentary committee on banking and credit. Wicksell's involvement with its work and that of its successor committees lasted until his death. He and Davidson were both appointed as experts to another parliamentary committee on taxation of income and property which remained active from 1918 to 1922. These assignments improved Wicksell's finances, for he was paid somewhat more than his pension for his work with these committees.

Among achievements attributable to Wicksell's and Davidson's collaboration in these councils was the adoption in 1916 of the 'gold exclusion policy' for the Bank of Sweden (to limit inflation the Bank was relieved of the obligation to issue currency at the pre-war mint

ratio to gold that had been turned in to it from Sweden's export surplus, and was given power to lower the price of gold in terms of currency). A second achievement was a thorough revision and improvement of the country's taxation of income and wealth.

In this decade, Wicksell also became a much sought-after adviser to young economists about their dissertations. At Lund he had only had three students in economics who took the intermediate graduate degree of *phil. lic.* under his guidance. In Stockholm, as a very active member of the Swedish Economics Association, and an indefatigable participant in the Economy Club, its inner circle of economists (as distinct from such members as bankers and industrialists), he had easy access to the club members' graduate students. He was made president of that club, 1917–22. It was a source of satisfaction for him to be sought out to share in the problems of the young men.

Thus his teaching did not stop with his retirement, for Emil Sommarin, Erik Lindahl, the brothers Gustav and Johan Ákerman, Bertil Ohlin, and probably others such as Palander, Lundberg and Hammarskjöld consulted him about their dissertations in addition to benefiting from studying his treatises.

In the 1930s these persons, self-confessed 'Wicksellians', formed the core of the 'Stockholm School of economists'. However, he remained estranged from Gustav Cassel, the third of Sweden's leading economists in the 1920s. This had nothing to do with Cassel's competition with him for the position at Lund in 1900; it was due to Cassel's character. Wicksell found him to be intellectually arrogant, rarely acknowledging the contributions or predecessors whose works he was using, and acting as if economics had been in its infancy prior to Cassel. Wicksell found some of his work to be superficial and his interpretations of several points in economic theory to be misleading. This he expressed clearly in his rather severe review in 1919, of Cassel's magnum opus, *The Theory of Social Economy* (Wicksell 1919b).

After that the breach between them was complete. Cassel never replied to Wicksell's review. As a result, Cassel remained an outsider to 'the Stockholm School', although he was the mentor of one of its leading members, Gunnar Myrdal.

In spring 1926, Wicksell was working on an article 'Zur Zinstheorie' (On the theory of interest) for a Festschrift for Friedrich von Wieser, when he fell ill with a stomach disorder which was further complicated by pneumonia. He died on 2 May 1926, at age 74.

His widow Anna, then a delegate from Sweden to the League of Nations, survived him until 1928. Their eldest son, Sven, became a professor of statistics at Lund University and died in 1939. Their young son, Finn, died in an accident in 1913 at the age of 19, when a medical student at Lund University.

Knut Wicksell would doubtless have objected to the elaborate funeral that was arranged for him, evidently with his widow's consent. Throughout life he had steadfastly rejected as meaningless and offensive to his sense of rationality all pomp and circumstance, academic formalities along with marriage ceremonies, baptism and confirmation for his children.

Contributions to Economics

In his own lifetime Wicksell did not receive much recognition for his creative work, not even in Scandinavia. It was not until the 1930s, when at the initiative of R.F. Kahn and J.M. Keynes, *Geldzins as Interest and Prices* and *Vorlesungen* as *Lectures on Political Economy I and II* were translated, that economists generally heard of Wicksell. Yet is it clear that his stature in the annals of economics grew steadily after his death. In summary form, his main contributions were these.

In *Value, Capital, and Rent* he performed a remarkable labour of synthesis. He adopted the marginal utility and marginal productivity theory of value of Jevons, Menger and Marshall, added to it the Böhm-Bawerk analysis of capital, and fused the result in a Walrasian comparative static general equilibrium framework. In this process he became a founder of the marginal productivity (product exhausting) theory of distribution shortly ahead of Wicksteed. In his *Studies in the Theory of*

Public Finance he pioneered a marginal utility approach to the public sector, synthesizing the benefit and ability principles of taxation, and urging that services of public sector enterprises and natural monopolies be provided on a marginal cost pricing basis.

In *Lectures I* he completed the restructuring, begun in *Value, Capital and Rent,* of Böhm-Bawerk's theory of capital and interest. He reduced Böhm's trinitarian 'grounds' for interest to the simpler, more realistic explanation as the marginal productivity of waiting. He relaxed Böhm's quantification of capital as an average period of production by a concept of capital as the time structure of inputs invested for various terms in production. He showed that this structure was capable of change in at least two dimensions, width and height. He endeavoured with partial success (on problems still unresolved about 'Wicksell effects' and 'switching of techniques') to develop a theory of the modes of change of this time structure of production, how it changes and interacts with variations in wages, rent, and interest, in conditions both of capital accumulation and technological change. He extended his treatment of these relationships from comparative static to dynamic analysis, using clear mathematical models for this purpose.

The greatest contribution to monetary analysis, both in terms of novelty back in 1898 and 1906, and in terms of eventual influence by fortifying the related analysis, independently worked up by J.M. Keynes three decades later in his *Treatise on Money* (1930), was Wicksell's work on monetary theory in *Interest and Prices* and especially in *Lectures on Political Economy II*.

Wicksell was a pioneer of applying an aggregate demand-supply approach with emphasis on the relations between investment and saving, to explain variations in value of money or fluctuations in prices: 'Any theory of money worthy of the name must be able to show why pecuniary demand for goods exceeds or falls short of the supply of goods in given conditions' *(Lectures II,* p. 160).

Most versions of the quantity theory had the price level varying directly and proportionately to changes in the quantity of money. In that theory there was no link from the elasticity, as affected by bank credit, and quantity of money with individual income dispositions and entrepreneurial production decisions. Wicksell provided such a link in his hypothesis that in the absence of certain disturbances over which central banks have no control (such as large influx or efflux of gold, internal cash drains, large government deficits financed by loans from the central bank, fiat issues of inconvertible currencies, sudden and large changes in productivity or supply of goods), price level fluctuations were due to a persistent divergence between the bank rate or market rate of interest and the real rate, defined as the expected rate of return on newly produced capital goods.

> The fluctuations of commodity prices, which are *not* due to a change in gold production [a gold standard currency is assumed here] . . . have another cause . . . changes in the real rate of interest. . . [to which] . . . the loan-rate does not adapt itself quickly enough. *(Lectures II,* p. 205)

Thus Wicksell's analysis showed, contrary to that of the simple quantity theory, that it was the quantity of money that adapted itself to the movement of the price level, and in doing so affected the distribution of income and the dispositions to invest and save in the process.

In his analysis monetary equilibrium and stability of prices required the simultaneous fulfilment of the conditions that: (i) the money rate of interest correspond to the real rate; (ii) at that money rate demand for loans for investment and for cash for real balances equal the supply of savings by individuals and business enterprises; and (iii) that interest rate must be neutral in its effect on prices. Then: '. . . equilibrium must *ipso facto* obtain – if not disturbed by other causes–in the market for goods and services, so that wages and prices will remain unchanged' *(Lectures II,* p. 193).

The consistency and compatibility of Wicksell's three criteria for monetary equilibrium and a critique of them in conditions of changing productivity by Davidson were given a thorough exegesis and analysis in the later 1920s and early 1930s by Lindahl, Myrdal and Ohlin. Their work combined with the efforts of younger colleagues

such as Lundberg, Hammarskjold and Svennilson greatly expanded the heritage of Wicksellian economic theory and gave rise to the doctrines associated with the Stockholm School of economics.

See Also

► Stockholm School

Selected Works

1890a. Tomme maver–og fulde magasiner [Empty stomachs and full stores]. *Samtiden* [Contemporary Times, a Norwegian periodical] 1: 245–247, 293–320.

1890b. Überproduction oder Überbevölkerung [Excess production or excess population]. *Zeitschrift für die gesamten Staatswissenschaften* 46.

1892. Kapitalzins und Arbeitslohn [Interest and wages]. *Jahrbücher für Nationalökonomie* 59: 552–574.

1893. *Über Wert, Kapital, und Rente.* Jena: G. Fischer. Trans. S.H. Frowein as *Value, capital and rent.* London: Allen & Unwin, 1954. Reprinted, New York: Augustus M. Kelley, 1970.

1894. Våra skatter: hvilka betalar dem, och hvilka border betala? [Our taxes: who pays them, and who ought to pay them?] Stockholm. This was one of Wicksell's early and very popular tracts, written under the pseudonym of Sven Trygg. It provides non-technical background and may be viewed as an introduction to Wicksell (1896).

1896. *Finanztheoretische Untersuchungen nebst Darstellung und Kritik des Steurewesens Schwedens.* Jena: G. Fischer. Pages iv–vi, 76–87 and 101–59 trans. J.M. Buchanan as Chapter 6 of *Classics of public finance,* ed. R.A. Musgrave and A.T. Peacock. London: Macmillan, 1958; 2nd ed., 1967. This treatise has only been translated in part; the untranslated sections deal with an historical sketch of the development of Sweden's system of taxation from the early 16th century up to the 1890s.

1898. *Geldzins und Güterpreise bestimmenden Ursachen.* Jena: G. Fischer. Trans. R.F. Kahn as *Interest and prices. A study of the causes regulating the value of money.* London: Macmillan, 1936.

1900. Om gränsproduktiviteten såsom grundval for den nationalekonomiska fördelningen. *Ekonomisk Tidskrift* 2: 305–337. Trans. as 'Marginal productivity as the basis for distribution in economics' in *Selected papers on economic theory by Knut Wicksell,* ed. E. Lindahl. London: Allen & Unwin, 1958.

1901. The theory of exchange in its final form. In *Föreläsningar i nationalekonomi. Häft I.* Stockholm/Lund: Fritzes, Berlingska. The 3rd Swedish edn (1928) of this volume translated by E. Classen as *Lectures on political economy.* Vol. 1: *General theory.* London: Routledge & Kegan Paul, 1934.

1902. Till fördelningsproblemet. *Ekonomisk Tidskrift* 4: 424–433. Translated as 'On the problem of distribution', in *Selected papers on economic theory by Knut Wicksell,* ed. E. Lindahl, 1958.

1904. Mål och medel i nationalekonomien. *Ekonomisk Tidskrift* 6: 457–474. Translated as 'Ends and means in economics' in *Selected papers on economic theory by Knut Wicksell,* ed. E. Lindahl, 1958.

1905. *Socialiststaten och nutidssamhället* [The socialist state and contemporary society]. Stockholm. A popular tract, 40 pp.

1906. *Föreläsningar i nationalekonomi. Haft II: Om penningar och kredit.* Stockholm/Lund: Fritzes, Berlingska. The 3rd Swedish edn (1929) trans. E. Classen, ed. L. Robbins, as *Lectures on political economy.* Vol. 2: *Money.* London: Routledge & Kegan Paul, 1935; reprinted 1946.

1907. Krisernas gåta. *Statsekonomisk Tidskrift.* Oslo. Trans. by C.G. Uhr as 'The enigma of business cycles' in *International economic papers* No. 3 (1953), 58–74. This article expands on and lifts to a higher level of analysis the rather primitive treatment of business cycles presented in Wicksell (1890a and 1890b) and in Wicksell's brief 'Note on trade

cycles and crises' in *Lectures on political economy,* vol. 2.

1910. *Laran om Befolkningen, dess Sammansättning och Förändringar* [Theory of population, its composition, and modes of change]. Stockholm: A. Bonniers, 92 pp. A popular tract written while Wicksell served a two-month jail sentence in southern Sweden for disturbing the religious peace by public blasphemy in a public lecture he had given in Stockholm in 1908.

1919a. Växelkursernas gåta. *Ekonomisk Tidskrift* 21: 87–103. Trans. as 'The riddle of the foreign exchanges' in *Selected papers on economic theory by Knut Wicksell,* ed. E. Lindahl, 1958.

1919b. Professor Cassels ekonomiska system. *Ekonomisk Tidskrift* 21: 195–226. This article, highly critical of several of Cassel's interpretations and formulations of economic theory, has been translated and added as Appendix 1 in *Lectures on political economy,* vol. 1, 1934.

1923. Realkapital och kapitalränta. *Ekonomisk Tidskrift* 25: 145–80. A review and a mathematical elucidation of the analysis in G. Åkerman's doctoral dissertation *Realkapital und Kapitalzins,* Lund, 1923. It has been translated and added to *Lectures on Political Economy,* vol. 1, as Appendix 2, 'Real Capital and Interest'. Among other things, this article features demonstrations of both the 'Wicksell effect' and the reversal of that 'effect' in an effort to determine the relationship between the optimum durability of fixed capital and the rate of interest. As such, this article, long after Wicksell's death, has played an important role in the controversy over capital theory between the Cambridge (England) and Cambridge (Massachusetts) economists.

1925. Valutaspörsmålet i de skandinaviska länderna. *Ekonomisk Tidskrift* 27: 103–25. This article has been translated and added to *Interest and Prices* as an Appendix as 'The Monetary Problem of the Scandinavian Countries'. It represents the several qualifications Wicksell was moved to add to his norm of price level stabilization for monetary policy, qualifications both to meet Davidson's criticism of this norm and to incorporate lessons from the monetary experiences and upheavals of the First World War and its aftermath in the early 1920s. The qualifications were introduced to make allowance for significant increases in productivity, and for something like its opposite, severe commodity shortages due to wartime blockades, crop failure and also for significant issues of fiat money by governments running large deficit budgets, and so on.

A full-scale bibliography of all of Wicksell's published writings is now available. Its author is Dr. E.D. Knudtson, who has written *Knut Wicksells Tryckta Skrifter 1868–1950* [The published writings of Knut Wicksell, 1868–1950] edited by T. Hedlund-Nyström, and issued in the series Acta Universitatis Lundensis, Section I, Theologia-Juridica-et-Humaniora, No. 25, and published by the C.W.K. Gleerup Publishing House, Lund, Sweden, 1976. This bibliography runs to slightly more than 100 pages and accounts for 889 titles or items dating from Wicksell's student days in the later 1860s through his entire career, inclusive of his many popular articles for Sweden's leading newspapers, and beyond, to include also listings of the translations of his major works and reviews of these translations, which appeared between the decade or two after Wicksell's death.

Bibliography

Blaug, M. 1985. *Economic theory in retrospect*, 4th ed. Cambridge: Cambridge University Press. Chapter 12, especially 'Reader's Guide to *Lectures on Political*.

Economy, vol. I', pp. 546–565: see also some of the earlier part of this chapter, pp. 498–519. Chapter 15, 'Reader's Guide to Wicksell's *Lectures,* vol. 2', pp. 646–51, and also in the earlier part of this chapter, pp. 637–45.

Gårdlund, T. 1958. *The life of Knut Wicksell*. Stockholm: Almqvist & Wiksell. The original Swedish version of this biography; *Kunt Wicksell – Rebell i Det Nya Riket* [Rebel in a new realm], Stockholm, 1956, is somewhat longer than the English version, largely because it contains, but the English omits, notes on source materials and valuable bibliographical data, pp. 374–405.

Hedlund-Nyström, T., and L. Jonung. 1986. Knut Wicksell's opublicerade manuscript. 2 vols. Mimeo, Department of Economics, University of Lund.

Jonung, L. 1987. Knut Wicksell's unpublished manuscripts. Mimeo, Department of Economics, University of Lund.
Uhr, C.G. 1951. Knut Wicksell – A centennial evaluation. *American Economic Review* 41: 829–860.
Uhr, C.G. 1960. *Economic doctrines of Knut Wicksell*. Berkeley: University of California Press.

Wicksell's Theory of Capital

Massimo Pivetti

Wicksell first developed his real theory of capital on 'the purely imaginary assumption' that the phenomena of capital and interest could take place without the intervention of money or credit; he then endeavoured to bring to light the modifications that are called for by the appearance of money, and by so doing he laid the foundation of this century's dominant approach to money and real magnitudes. Wicksell's theory can actually be said to have established the basis of mainstream long-period analysis of the economy, with its explanation in real terms of the equilibrium rate of interest and the conception of money as a factor that may be important to the gravitation of the economy towards its equilibrium position but not as a determinant of that position.

Wicksell's general equilibrium – what he calls 'The Theory of Exchange Value in its Final Form' (1901, p. 196) – consists of a system of equations by which relative prices are determined simultaneously with normal outputs, factor uses and the equilibrium prices of factor services (that is, distribution), on the basis of given consumer tastes, technical conditions of production and factor endowments. 'Capital' enters *twice* into the picture: first, when the quantity produced in the economy of each final article is expressed as a function of all the quantities of factors employed, according to given 'production functions' reflecting the given technical conditions; secondly, in the relations expressing the condition that the supply of each factor of production annually available in the economy must be equal to the quantity of it annually employed (i.e. demanded). On the basis of these conditions, the equilibrium prices of factor services depend on their relative scarcities; the equilibrium rate of interest, in particular, depends on the relative scarcity of the whole available capital and is the same on all capital (Wicksell 1901, pp. 144–6). Wicksell's system thus depicts a 'long-run' equilibrium, which ultimately reflects the idea, common to the original versions of the marginal theory, that the competitive tendency towards a uniform rate of interest (profit) would deprive of any significance, as centres of gravitation of the economy, quantities and prices determined for situations in which each particular capital-good gave a different rate of return over its cost. (The same idea explains Wicksell's attention being focused throughout his main writings on *circulating* capital, for which the equilibrium condition represented by a uniform rate of return tends rapidly to impose itself through changes in the proportions amongst the different kinds of capital-goods annually employed in the economy. As for the treatment of durable or fixed capital, see Wicksell 1923.)

The important point to be noticed about Wicksell's 'production functions' is that the capitalist element is expressed in them not by means of *value* magnitudes but in 'technical units'. He was fully aware that the partial derivatives of any such function in which 'capital' appears in value terms can be of no use for determining the 'productive contribution' of the different productive factors, and hence distribution. (An increase in the value of capital may simply reflect a rise of wages and rent, possibly without causing any change in the magnitude of the return; the additional product of the new capital may thus be nil, but this would give no information at all about the new level of the rate of interest: see Wicksell 1893, pp. 25, 115–19; 1901, p. 148.)

In *Value, Capital and Rent* Wicksell used Böhm-Bawerk's 'average period of production' (Böhm-Bawerk 1889, vol. II, bk. II, ch. II) so that the role of capital was seen as making possible the introduction of a longer period of time between the beginning and the conclusion of the process of production 'and consequently the

adoption of a more roundabout method of production than would be possible if production were less strong in capital or totally devoid of capital'. He maintained, accordingly, that the greater the amount of capital employed, 'that is to say, the lengthier the average period of production that can be applied, the greater will be the annual production of finished consumption goods, provided the same number of workers and the same area of the country are involved' (1893, p. 116).

Wicksell realized, however, that the average period of production made it necessary to have recourse to calculation with simple interest (1893, pp. 125–6; 1901, p. 205 and Preface to the 2nd edn); so in the *Lectures* (1901), while still adhering to Böhm-Bawerk's view of the role of capital, Wicksell moved to a conception of capital in the production functions as a *complex* of variables: dated quantities of labour and land (or 'saved-up labour and saved-up land' as he called them to indicate that instead of being quantities of current labour and land *directly* employed in the production of consumption goods, they are employed in the production of capital goods). Wages and rents actually paid to those quantities of labour and land remain 'invested' in production from the moment they are paid until the conclusion of the process of production of the consumption good concerned; on the other hand, their marginal productivities are greater than those of current labour and land directly employed in production – the idea being that the productivity of original factors becomes greater if they are employed for distant ends than if they are employed in the immediate production of consumption goods: this difference in productivity constitutes the very source of interest (1901, p. 154).

Now, for the *rate* of interest to be same on all kinds of investment (in labour-capital or in land-capital, for a single year or for a period of years), the marginal productivities of the dated quantities of original factors – that is, the partial derivatives of the production function with respect to each of the variables included in it – must stand in a certain relation to each other, 'corresponding to that which exists in a calculation with compound interest' (p. 160). Full equilibrium determination then entails that the dated quantities of labour and land appearing in the production functions cannot be taken as given, but must be included amongst the unknowns of the system (pp. 203–5).

We can sum up the above by saying that the notion of marginal productivity is never applied by Wicksell directly to capital or capital goods; it is applied (in his chief work) to dated quantities of the two original 'factors of production'. Each commodity is seen as ultimately resolving itself into labour and land employed in different years – current labour and land, and 'saved-up' labour and land; they are remunerated according to their marginal productivities and are employed in the proportions demanded by the equilibrium condition of a uniform rate of interest. The equilibrium level of the rate of interest ultimately reflects the relative scarcity of saved-up original factors:

> the marginal productivity of the latter is greater, simply because current labour and land exist in relative abundance for the purposes for which they can be employed, whilst saved-up labour and land are not adequate in the same degree for the many purposes in which they have an advantage. This again is to be explained by the circumstances which limit the accumulation of capital (p. 155).

The explanation of the equilibrium rate of interest by the scarcity of capital, and as the reward for 'waiting', is one and the same thing as the conception of the rate of interest as the variable that brings to equality the supply of and the demand for capital. To this equality we now turn our attention.

In the relations expressing the equality between the supply of and the demand for each factor of production, together with the total quantities of labour and land, the total quantity of capital annually available in the economy is taken as given; it is a single magnitude, so that what Wicksell actually takes as given to solve the system is the total *exchange value* (measured in terms of one of the final products) of the capital available in the economy (pp. 204–5). A value magnitude is thus included amongst the determinants of distribution and prices. In criticizing Walras for having taken the physical quantities of the different kinds of capital goods as given, Wicksell argued that we need 'a unified treatment of the role of capital in production . . . in order to

calculate the rate of interest, which in equilibrium is the same on all capital' (p. 149). Accordingly, the quantity of capital available in the economy is conceived in his system as a single magnitude, a value magnitude taken as given, whilst, as we saw above, its physical composition, the relative quantities of its different technical constituents, is left free to change during the process of adjustment to equilibrium in order to satisfy the condition of a uniform rate of interest.

If the quantity of capital available in the economy is increased by 'real, productive, saving' (i.e. 'by restricting or postponing consumption'), then, *ceteris paribus,* the equilibrium rate of interest must fall. At the old rate the supply of capital now exceeds the quantity of it annually employed in the various industries; competition amongst capitalists presses the rate of interest downwards, thereby causing more roundabout processes, which were previously unremunerative, to become profitable. In Wicksell's view, not only does the process of production of each consumption good tend to become more 'capitalistic', through increases in 'saved-up' labour and land relative to current labour and land used in the course of a year and the introduction of 'longer-dated' investments, but also the composition of final demand and output tends to change in favour of more 'capitalistic' consumption goods, through the relative cheapening of such goods brought about by the fall in the rate of interest. Following, therefore, an increase in the supply of capital, substitution amongst alternative methods of production and amongst alternative consumption goods would ensure a new equality between supply of and demand for capital at a new lower level of the rate of interest. It may conveniently be added that if only circulating capital is taken into consideration – as Wicksell actually did by centering his theory of capital upon the case of capital goods that last only one year in an economy where production takes place in yearly cycles – then there is no need to distinguish, in the determination of interest, between supply of and demand for capital as a stock and as an annual flow; one may simply refer to the equilibrium rate of interest as being determined by supply and demand for gross saving.

This explanation of the 'real capital rate', with the essential role played in it by the interest elasticity of demand for saving, constitutes the basis of Wicksell's theory of money and prices (1898a, b, 1906). We shall here refer to its more mature version, contained in volume II of the *Lectures* (1906).

In Wicksell's opinion, 'any theory of money worthy of the name must be able to show how and why the monetary or pecuniary demand for goods exceeds or falls short of the supply of goods in given conditions'. He contended that the advocates of the Quantity Theory, in postulating the price-level as an increasing function of the quantity of money, failed to show 'why such a change of price must always follow a change in the quantity of money and to describe what happens' (1906, p. 160). In the solution he put forward, the primary cause of price fluctuations is singled out as the difference between the actual money or loan rate and the normal or natural real rate of interest, determined by the scarcity of capital (saving).

As we saw above, *ceteris paribus* a lowering of the real rate unconditionally demands increased saving. The same applies to a lowering of the loan rate in the case of 'simple credit between man and man'; the loan market would directly reflect in such a case the supply of and demand for saving, so that there would be an immediate connection between the money rate and the real capital rate. Changes in the loan rate would take place simultaneously and uniformly with corresponding changes in the real rate, with the result that no change in the level of commodity prices could occur. Things are different when the activity of the banks is taken into consideration: banks 'possess a fund for loans which is always elastic and, on certain assumptions [i.e. with a pure credit system], inexhaustible', with the consequence that the immediate connection between the money rate and the real natural rate disappears. 'In our complex monetary system', says Wicksell, 'there exists no other connection between the two than the *variations in commodity prices* caused by the difference between them' (1906, pp 194, 206).

Thus, starting from an equilibrium situation and no changes occurring in the circumstances upon which the real natural rate depends, a reduction of the rate of interest on the part of the banks

will lead to an increase in monetary demand: owing to the increased demand for loan capital and the expansion of credit, on the one hand, and to a reduced supply of saving, on the other, an excess of investment spending over saving decisions will arise. Since the normal or equilibrium situation of the economy is characterized by the full employment of all productive factors, the increased monetary spending will result in a rise in prices, both of production and of consumption goods. As the parallel rise in money prices and incomes tends to leave the real capital rate unaffected, at the new higher price level an excess of investment spending over saving decisions will present itself again – so that the inflationary process is bound to continue as long as the money rate is kept below the real rate. The opposite would occur if banks maintained the rate of interest above its natural level. In both cases, in order to re-establish monetary equilibrium – the stability of the price level – banks would have to bring the money rate of interest back to the level of the 'real natural rate'. The conclusion then is that by virtue of the 'connecting link' of price-movements, the money rate will gravitate towards the real rate, even if such a gravitation process will not be of an automatic-spontaneous nature. (We may add that persistent full employment does not seem to be essential to the Wicksellian notion of a non-automatic gravitation of the actual money rate towards the normal natural rate. If, in the face of a money rate of interest that is higher than the natural rate, rigid money wages are assumed, then the role of 'connecting link' between the two rates might be played not only by reductions in prices but also by the fall in employment.)

So, in this picture it is maintained that a low rate of interest causes prices to rise, and vice versa. But Wicksell recognizes that in actual experience rising prices very rarely coincide with low or falling interest rates, and that the opposite is the general rule. He argues that this 'apparently crushing objection' to his theory ('an objection which the members of the Tooke School have triumphantly produced at every opportunity as a support for their theory') is indeed perfectly consistent with his view of the influence of the rate of interest on prices: instead of assuming a lowering of the rate of interest by the banks, other things being equal, one has simply to make the more realistic assumption that the difference between the two rates arises because the natural rate rises or falls whilst the money rate remains unchanged and only tardily follows it. The *primum movens*, that is to say, generally consists in changes in the natural real rate of interest: a rise (fall) in the natural rate will result in a rise (fall) in prices which, in its turn, will sooner or later force up (down) the money rate (1906, pp. 202–8; see Keynes's identical explanation in *A Treatise on Money,* 1930, vol. I, p. 196n. and vol. II, pp. 198, 203).

The critique of the marginalist notion of capital which was stimulated by the work of Piero Sraffa (1960) applies also to Wicksell's theory. Amongst the numerous relevant contributions, we shall recall here the careful critical analysis of Wicksell's theory of capital contributed by Garegnani (1960, chs. 4–6, see also Garegnani 1970), and the symposium on capital theory in the *Quarterly Journal of Economics* (1966) with the contributions by Pasinetti, Samuelson, Morishima, Burmeister and others.

As we saw above, in Wicksell's system the quantity of capital annually available in the economy is taken as given in terms of a single magnitude, which is thus included amongst the determinants of general equilibrium. At the same time, a decreasing demand function for capital (saving) is postulated, based on the 'substitution' principle – the principle according to which a fall in the rate of interest cheapens the more capital-intensive processes of production relative to the others, thereby raising the proportion of capital to the other productive factors in the economy. Such a demand function is essential to the idea that, *ceteris paribus,* increased saving will result in a reduced natural rate of interest, hence to the explanation of interest by the scarcity of capital.

Both these aspects of the marginalist analysis of capital have been found faulty on logical grounds, the main ingredients of the critique having been provided by Sraffa (1960). By studying the movement of relative prices consequent upon changes in distribution, he found that, in the face of unchanged methods of production, *reversals* in the direction of that movement may occur – a

phenomenon which 'cannot be reconciled with *any* notion of capital as a measurable quantity independent of distribution and prices' (1960, p. 38); that is, of the very unknowns that the quantity of capital available in the economy should contribute to determine. (No such reversals could possibly occur if a single magnitude existed which was both independent of distribution and prices *and* representative of capital. Böhm-Bawerk's average period of production, for example, is independent of distribution and prices, but is not representative of the quantity of capital: if it were, then, assuming an average period of production of commodity *A* greater than that of commodity *B*, *pa* would *continuously* rise relative to *pb* with the rising of the rate of interest, contrary to what is shown by Sraffa (para. 48).) Moreover, the reversals in the direction of the movement of relative prices and the analogous phenomenon of 'reswitching' of methods of production (ch. XII) entail that no demand function for capital (saving) can be deduced from the existence of alternative methods of production and alternative consumption goods, except in very restrictive hypotheses.

We pointed out above the crucial role played in Wicksell's theory by the concept of a natural real rate of interest. The same concept plays a significant role also in Keynes's writings. This is clearly so in the *Treatise;* but also in the *General Theory,* notwithstanding the author's statement that he no longer regards the concept as 'a most promising idea' (Keynes 1936, p. 243), yet the 'natural rate' is still there, as the rate of interest that would ensure equality between full employment saving and investment decisions. Keynes's underemployment equilibrium is ultimately the result of the presence in the economic system of factors that hinder the possibility of bringing the actual rate of interest down to its 'natural' or full employment level – it is the result, in other words, of a limited flexibility of the money rate of interest. If one takes into account that also in Wicksell there is no automatic gravitation of the money rate towards the level of the natural real rate (banking policy having to perform the task; see above), then the difference between the two authors will not appear so marked: they both share, in particular, the idea of an inverse relation between the rate of interest and investment decisions, whilst the contrast of opinion is essentially centred upon the degree of (non-automatic) flexibility of the rate of interest in the face of discrepancies between full employment saving and investment decisions. We believe that largely in the light of such a comparison the thesis was successfully laid down that, far from constituting *the* general theory, 'The General Theory of Employment is the Economics of Depression' (Hicks 1937, p. 154). So our point here is that a better knowledge of Wicksell's work would have greatly facilitated the singling out of the traditional premises in *The General Theory* that aided the subsequent 'neoclassical synthesis', thereby helping to realize the importance of the critique of the marginal theory of capital and interest for establishing Keynes's principle of effective demand on firmer ground.

See Also

- ▶ Accumulation of Capital
- ▶ Austrian Economics
- ▶ Böhm-Bawerk, Eugen von (1851–1914)

Bibliography

Böhm-Bawerk, E. von. 1889. *Capital and interest: A critical history of economic theory.* Trans. W. Smart. New York: Stechert, 1932.

Burmeister, E. 1976. Real Wicksell effects and regular economies. In *Essays in modern capital theory*, ed. M. Brown, K. Sato, and P. Zarembka. Amsterdam: North-Holland.

Garegnani, P. 1960. *Il capitale nelle teorie della distribuzione*. Milan: Giuffrè.

Garegnani, P. 1970. Heterogeneous capital, the production function and the theory of distribution. *Review of Economic Studies* 37(3): 407–436.

Hicks, J.R. 1937. Mr. Keynes and the 'classics': A suggested interpretation. *Econometrica* 5(2): 157–159.

Keynes, J.M. 1930. *A treatise on money.* London: Macmillan.

Keynes, J.M. 1936. *The general theory of employment, interest and money.* London: Macmillan.

Pasinetti, L., D. Levhari, P.A. Samuelson, M. Morishima, M. Bruno, E. Burmeister, E. Sheshinski, and P. Garegnani. 1966. Paradoxes in capital theory: A symposium. *Quarterly Journal of Economics* 80(4): 503–583.

Sraffa, P. 1960. *Production of commodities by means of commodities: Prelude to a critique of economic theory.* Cambridge: Cambridge University Press.
Wicksell, K. 1893. *Über Wert, Kapital und Rente*. Jena: G. Fisher. Trans. by S. H. Frowein as *Value, capital and rent.* London: Allen & Unwin, 1954; reprinted. New York: Kelly, 1970.
Wicksell, K. 1898a. *Geldzins und Güterpreise bestimmenden Ursachen*. Jena: G Fischer. Trans. by R.F. *Kahn as interest and prices: A study of the causes regulating the value of money.* London: Macmillan, 1936.
Wicksell, K. 1898b. The influence of the rate of interest on commodity prices. In *Selected papers on economic theory*, ed. K. Wicksell. London: Allen & Unwin, 1958.
Wicksell, K. 1901. *Föreläsningar i nationalekonomi. Häft I.* Stockholm. Lund: Fritzes, Berlingska. The 3rd Swedish ed (1928) of this volume trans. by E. Classen, ed. L. Robbins, as *Lectures on political economy,* vol I: *General theory.* London: Routledge & Kegan Paul, 1934.
Wicksell, K. 1906. *Föreläsningar i nationalekonomi. Häft II: Om penningar och kredit.* Stockholm/Lund: Fritzes/Berlingska. The 3rd Swedish ed (1929) trans. by E. Classen, ed. L. Robbins, as *Lectures on political economy,* vol II: Money. London: Routledge & Kegan Paul, 1935.
Wicksell, K. 1923. Realkapital och kapitalränta. *Economisk Tidskrift* 21: 45–80. Trans. by S. Adler as *Real capital and interest* as Appendix 2 to Volume I of the *Lectures.* London: Routledge & Kegan Paul, 1934.

Wicksteed, Philip Henry (1844–1927)

Ian Steedman

Keywords

Abstract utility; Aristotle; Choice; Cost of production; Distribution theory; Economic man; Euler's Theorem; Exchange value; George, H.; Intensive rent; Interpersonal utility comparisons; Jevons, W. S.; Labour theory of value; Land nationalization; Land reform; Law of indifference; Law of one price; Marginal productivity theory; Marginal productivity theory of the interest rate; Marginal utility; Marginalism; Marx, K. H.; Opportunity costs; Orderings; Ophelemity; Preference ordering; Preferences; Price discrimination; Product exhaustion; Real cost doctrine; Reservation price; Revealed preference theory; Shaw, G. B.; Utility theory of value; Wicksteed, P. H.

JEL Classifications

B31

Wicksteed was born in October 1844 in Leeds, where his father, Charles Wicksteed, was a Unitarian minister. He died, at the age of 83, in March 1927, at Childrey in Berkshire. He attended Ruthin Grammar School in North Wales and then University College School, London, before studying at University College London (1861–1864) and at Manchester New College (1864–1867) in Gordon Square nearby. He received his Master's degree, with a gold medal for classics, in 1867. Wicksteed then became a Unitarian minister, first at Taunton in Somerset (1867–1869), then at Dukinfield, east of Manchester (1870–1874), and finally at Little Portland Street Chapel, London (1874–1897). He left the ministry in 1897 and thereafter earned his living by writing and lecturing. From 1887 to 1918 Wicksteed was a most active University Extension Lecturer, lecturing on Wordsworth, Dante, Greek tragedy, Aristotle and Aquinas – and economics. He never held a university post.

The great breadth of Wicksteed's intellectual activity was far from being confined to his Extension lecturing. He had a considerable linguistic talent; whilst a minister in Dukinfield, for example, he learned Dutch for the express purpose of translation into English of Oort and Hooykaas's *Bible for Young People* (six volumes, 1873–1879). And he completed a translation, with F.M. Cornford, of Aristotle's *Physics* only days before his death. Yet it was as a translator, expounder and interpreter of Dante that he became most widely known; his work as a Dante scholar, which extended over more than 40 years, included translations of and commentaries on the *Vita Nuova*, the *Convivio, De Monarchia* and the *Divina Commedia*. Combined with his theological and philosophical interests, this study of Dante led Wicksteed to Aquinas and thus to the writing of his *Dante* and *Aquinas* (1913) and his *Reactions between Dogma and Philosophy, illustrated from the Works of S. Thomas Aquinas* (1920). That a study of Aquinas' thought by a former Unitarian minister could be reviewed favourably in the

Blackfriars Review is perhaps an indication of the catholicity of Wicksteed's interests and capacities. Nor did those interests extend only to the past; for example, Wicksteed publicly defended the poetry and drama of Ibsen at a time when Ibsen's work was the object of considerable hostility in England. And Wicksteed's numerous contributions to the *Inquirer*, the Unitarian newspaper, over a span of some 50 years, relate not only to theological and literary matters but also to many economic and political issues.

While he had earlier been influenced by the thought of Comte and of Ruskin, Wicksteed's first direct contact with political economy took the form of reading Henry George's *Progress and Poverty*, of corresponding with George in 1882 and 1883 and of being a co-founder, in 1883, of the Land Reform Union, which supported George's lecture tour of England and Scotland in 1883–1885. (He continued to support some form of land nationalization long after this time.) It was probably late in 1882 that Wicksteed began to study the work of Jevons and thus to become 'Jevons's only disciple'. By early 1884, however, he was playing an active role in promulgating Jevonian theory in the Economic Circle, which met until 1888 or 1889 (to be followed by the Economic Club and the British Economic Association, later to become the Royal Economic Society). Wicksteed became a close friend of George Bernard Shaw and of Graham Wallas, and was well-informed about Fabian and other aspects of the 'social movements' of the 1880s and 1890s, but was generally an acute and sympathetic observer, rather than a direct participant in those movements. He was, however, a founder member, in 1891, of the Labour Church movement and continued to give that movement strong support even after other early supporters had withdrawn their active sympathy.

Wicksteed published three books in the field of economics. The first, *The Alphabet of Economic Science, Part I. Elements of the Theory of Value or Worth*, was published in 1888; the second, *An Essay on the Co-ordination of the Laws of Distribution*, was published in 1894, and the third work, *The Common Sense of Political Economy*, was first published in one volume in 1910; a second edition in two volumes, edited by L. Robbins and containing various papers and reviews by Wicksteed, was published in 1933.

Of Wicksteed's other writings in economics, the most important are probably his critique of *Das Kapital*, published in the socialist journal *To-Day* in 1884; his article on Jevons's *Theory of Political Economy* (1889); his various contributions to the first (1894) and second (1925) editions of *Palgrave's Dictionary of Political Economy*; and his 'Scope and Method of Political Economy' paper (1914), which originated as Wicksteed's Presidential Address to Section F of the British Association for the Advancement of Science in 1913. (All of these papers appear in the Robbins edition of the *Common Sense*.)

There are a few extant letters (Sturges 1975, p. 128) and some handwritten sermons and letters at Manchester College, Oxford, but Wicksteed wrote to a correspondent (J.M. Connell) that 'I have never kept careful records of my life and have next to no documents'. As to secondary material, the following may be consulted: Herford's full biography (1931); Robbins's editorial introduction (1933); the relevant chapters in Hutchison (1953) and Stigler (1941); Steedman's editorial introduction (1987); and the relevant entries in the *Encyclopaedia of the Social Sciences* (by H.E. Batson, vol. 15, 1935) and in the *International Encyclopedia of the Social Sciences* (by W.D. Grampp, vol. 16, 1968).

Wicksteed's first contribution to economic theory was his October 1884 critique of *Das Kapital*, volume 1. Resulting perhaps from a Fabian challenge within the Economic Circle, it was published in *To-Day*, which, in 1884, carried articles by many of the leading socialists of the time. Wicksteed's critique certainly converted George Bernard Shaw from the Marxian to the Jevonian theory of value and, since no effective reply was published, may have had a wider influence on the theory adopted by British socialists: some writers have regarded Böhm-Bawerk's later attack on the labour theory of value, of 1896, as inferior to that of Wicksteed. Displaying a firm grasp of many of the specific features of Marx's argument, Wicksteed was able to focus clearly on two central issues. Is the exchange value of ordinary commodities determined by labour time? And does

Marx's argument apply to 'labour force' (as Wicksteed called it)?

With respect to the first question, Wicksteed follows Marx in saying that if two commodities are exchanged they must simultaneously differ from one another, to motivate the exchange, and have something in common, to make them commensurable. But he then seizes on Marx's point that labour time only 'counts' when producing something useful and argues that it was merely arbitrary for Marx to assert that commodities have only 'abstract labour' in common. On the contrary, Wicksteed insists, all commodities have 'abstract *utility*, i.e., power of satisfying human desires' in common; moreover, this is just as true of exchangeable objects which are not freely reproducible. Thus, in a neat twist of the argument, he proposes '*abstract utility* as the measure of value'. Wicksteed argues, nevertheless, that for freely reproducible commodities equilibrium relative prices will *coincide* with relative labour costs – but this is not because labour quantities determine prices but because labour will be so allocated as to produce those quantities of the commodities which imply marginal utilities proportional to the given labour costs. For old masters, the products of monopolized industries, and so on, even this coincidence will not hold.

Turning to 'the value of labour-force', Wicksteed then observes that, in a non-slave society, labour is not allocated to the production of 'labour-force' under competitive pressures. He deduces that there is no reason to expect that the ratio of the money wage rate to the labour value of the necessary wage goods will be equal to the money price-embodied labour ratio for ordinary commodities. Consequently, he concludes, Marx has failed to show that 'surplus labour' is the source of profit. Neither George Bernard Shaw nor any other contributor to *To-Day*, or to the other British socialist periodicals of the period, provided a remotely effective reply to Wicksteed's argument.

The *Alphabet*

Wicksteed's *Alphabet of Economic Science*, of 1888, was dedicated to members of the Economic Circle who had 'met to discuss the principles set forth in these pages'. Both the subtitle of the volume and certain remarks in Wicksteed's Introduction suggested that there might be successor volumes but this proved not to be the case. Although the work received the approbation of both Edgeworth and Pareto, it did not find a wide audience, which is perhaps not surprising given that it was simultaneously introductory and somewhat mathematical. As in his other books, Wicksteed disclaimed originality but showed himself to be, at the very least, a most careful and detailed thinker and expositor; in the case of the *Alphabet* a great many vivid examples are used to reinforce the reader's firm grasp of marginal principles. (The book's only index is indeed an index of examples.) As in his earlier reply to G.B. Shaw, of 1885, and in the subsequent *Co-ordination of the Laws of Distribution*, of 1894, Wicksteed emphasized the importance of the mathematical expression of marginal economic theory.

For Wicksteed, the theory of value – or 'worth' – means essentially the theory of demand (the theory of supply he refers to as that of production – or 'making'). In both the discussion of 'individual worth' (pp. 1–67) and that of 'social worth' (pp. 68–138), stress is firmly laid on the distinction between total and marginal utility. (Wicksteed uses the latter term and avoids Jevons's 'final utility' and 'final degree of utility'.) While the analysis is based on utility rather than on choice or preference – and 'hedonistic value' is referred to (p. 54) – Wicksteed's later stress on choice between satisfactions which are rendered comparable at the margin is already foreshadowed in the *Alphabet.* It is suggested that all marginal utilities and disutilities, for an individual, may be measured in terms of the hedonistic value, to that individual, of foot-pounds of lifting work or perhaps of one hour of correcting examination papers. Although the exposition is elementary throughout the book, the careful reader will notice Wicksteed's remarks on indivisible commodities and marginal analysis, on the acquiring of preferences, on minimum perceived differences, on traditions and habits, on the desire to impress or to give to others, and on negative marginal (and even total) utilities.

Turning to 'social worth', Wicksteed asserts at once that interpersonal comparisons of utility are impossible; all that can be said is that the ratio of the marginal utilities of any two commodities is the same for any two individuals who possess some of each commodity. (Wicksteed gives a particularly clear account of why this proportionality rule does not hold for an individual whose possession of one or both of the commodities is zero.) Yet he is still ready to argue, on grounds of 'averages' and probabilities, that a more equal distribution of income will probably make the objective social scale of relative prices a more reliable guide to the relative social importance, at the margin, of the various commodities. Wicksteed then discusses the market demand curve, the law of indifference (that is, of one price) and various kinds of price discrimination.

As indicated above, Wicksteed considers that 'Strictly speaking [the allocation of productive resources] does not come within the scope of our present inquiry' (p. 109) but he nevertheless devotes pages 109–24 to the allocation of 'the labour (and other efforts or sacrifices, if there are any others) needful to production' (p. 109). As in the *To-Day* essay of 1884, he argues that the relative prices of freely reproducible commodities will, in equilibrium, be equal to their relative effort-and- sacrifice costs but that this is *not* because production costs give commodities their exchange value. Rather it is because resources are reallocated until the commodities are produced in those quantities for which the marginal utilities – which *are* the sources of exchange value – will be proportional to the constant costs. Given that Wicksteed argues here in terms of 'a unit of effort-and-sacrifice' or 'a unit of productive force' (p. 112 and n.), it is not surprising that no theory of distribution is offered or, indeed, even hinted at.

Co-ordination of the Laws of Distribution

Wicksteed's *QJE* article of the following year, 1889, nevertheless contained an important passage criticizing and extending Jevons's marginal productivity theory of the interest rate, and distribution theory became more prominent in Wicksteed's lectures in the following years. This development culminated with the publication, in 1894, of his famous *Essay on the Co-ordination of the Laws of Distribution.* A number of writers in the early 1890s began to extend the marginal theory of intensive rent into a more general theory of distribution but it was Wicksteed's *Essay* which most clarified the issues involved. He noted that the traditional exposition of intensive rent theory, in which varying amounts of 'capital-and-labour' were applied to a fixed amount of land, had two crucial properties. First, that the argument essentially concerned only the *proportions* between inputs, and not their absolute levels, and second that the whole argument was *reversible* – the logic is quite unchanged if varying amounts of land are applied to a fixed quantity of 'capital-and-labour'. It was thus a mere matter of historical accident, Wicksteed argued, that the conventional diagram made one factor return appear as a 'marginal product' and the other as a 'surplus'.

Having argued that it was in any case self-evident that a profit-maximizing entrepreneur would hire each input up to the point at which its marginal value product equalled its (given) price, Wicksteed set himself the task of demonstrating that marginal product pricing of all inputs would entail product exhaustion. (He did not show that there would be any objection in principle to a theory in which *one* return was determined residually – nor could he have done so.) This he did by a long and inelegant mathematical argument, which amounts to no more (and no less) than a proof of Euler's Theorem for homogeneous functions, in the two-variable case. (As was quickly pointed out by Flux in a review in the *Economic Journal* for June 1894; there is some evidence to suggest that Wicksteed was completely unaware of Euler's Theorem before reading Flux's review.) More interesting than the inelegance of Wicksteed's proof, however, is that he was not satisfied with the argument, for while he considered it to be a 'truism' that there are constant returns to scale in physical production, he insisted that there might well not be constant returns in terms of revenue. Even if such 'commercial' factors as 'goodwill', 'travelling' and

'notoriety' could be increased in the same proportion as all the inputs to physical production, he argued, total revenue might increase in a smaller proportion. Wicksteed was thus led first to consider a monopolist (and to present quite explicitly the marginal revenue formula – already known to Cournot – of the imperfect competition theory of some 40 years later) and then to show how, as the number of firms in an industry becomes ever larger, the product exhaustion theorem will become 'virtually' correct. In his later review of Pareto's *Manuale* (1906) and in *The Common Sense* (1910, p. 373, n. 1) Wicksteed appeared to withdraw the sixth section of the *Essay* dealing with product exhaustion in the presence of monopoly, and so on (although not the *Essay* as a whole) but there has been considerable discussion of just how that apparent withdrawal ought to be interpreted.

Wicksteed's *Essay* constituted a major contribution to marginal productivity theory, by raising and discussing the product exhaustion question and by setting the theory very firmly in a multi-product, multi-input setting. (The practice of treating capital, or 'capital-and-labour', as a single sum of value is sharply criticized.) It is to be clearly noted, however, that the *Essay* presented *partial* equilibrium analysis throughout; Wicksteed always takes input prices as given and, contrary to some commentators, he never asserts that input supplies are exogenously determined. The *Essay* is a major text in partial analysis; it does not present a general equilibrium argument.

The *Common Sense*

From 1894 to 1910 Wicksteed published very little in the field of economics but in 1906 he was ready to begin work on his magnum opus *The Common Sense* of *Political Economy*, published in 1910. In this 700-page book, he sought to expound in minute detail the consequences of 'the revolution that has taken place' (p. 2) in economic theory. Disclaiming originality yet again, as he had done in 1888 and in 1894, and making very few *explicit* references to the work of others, Wicksteed presented a consistently subjective approach to all aspects of economic life. (Just five years earlier, in the *Economic Journal*, 1905, p. 435, he had written that 'The school of economists of which Professor Marshall is the illustrious head may be regarded from the point of view of the thorough-going Jevonian as a school of apologists.') Ranging from behaviour at the dining table to the significance of the division of labour in an advanced society, Wicksteed argued that attention to *selection between alternatives* was the key to understanding *all* aspects of allocation – whether of bread, of bricks, of friendship, of charity, of labour time or of prayers. Indeed, he even saw an intimate connection between careful marginal allocations and 'the law formulated by Aristotle with reference to virtue', that of the mean. The following discussion of Wicksteed's long, immensely detailed and occasionally prolix work will have to centre on his positive contributions and no reference will be made to weaker parts of his analysis (for example, that on increasing and diminishing returns in Book II, Chapter 5) or to his discussion of distribution theory, already referred to above in relation to the *Essay* of 1894. (Wicksteed's famous 'Scope and Method' paper, of 1914, presents an incisive epitome of the central themes of the *Common Sense* and may serve as an introduction to it.)

Wicksteed's analysis of choice, in the *Common Sense*, is firmly based on the concept of a scale of preferences, diminishing marginal significance and equivalence at the margin; it has been freed from the notions of utility and marginal utility as quantities, which are still evident in the earlier *Alphabet*. Moreover, while there is some room for doubt, in the *Alphabet*, whether the 'marginal utility' of a commodity depends only on the quantity of that commodity or on the quantities of all the commodities possessed, it is completely clear, in the *Common Sense*, that the 'marginal significance' of a quantity of a particular commodity depends on all the quantities in question. Indeed it depends not only on all those quantities but on all the circumstances of the choosing individual, for Wicksteed is insistent throughout that *all* objects of choice, and not just marketable commodities, have a bearing on each choice. The

principles at work in the allocation of money between potatoes and milk are the same as those involved in the allocation of time between friendship and prayer: 'whatever our definition of Economics and the economic life may be, the laws which they exhibit and obey are not peculiar to themselves, but are laws of life in its widest extent' (p. 160). Wicksteed's firm refusal to draw boundaries is more readily understood when account is taken of his conviction that 'these things, of which money gives us command, are, strictly speaking, never the ultimate objects of deliberate desire at all . . . as soon as we deliberately desire possession of any external object, it is because of the experiences or the mental states and habits which it is expected to produce or to avert' (p. 152). In modern terms, the underlying preference ordering is over mental experiences, not over commodities, and there is no reason to expect that 'economic' choices will fall under different principles than do 'other' choices.

The individual's preference ordering, Wicksteed argues, will be complete but will not always be consistent (transitive), although reflection will increase its consistency. The ordering often will not be, and will not need to be, fully present to the agent's consciousness. Apparently 'irrational' behaviour based on impulse, habit or tradition certainly occurs but does not undermine the fundamental principles of rational behaviour; 'Habit or impulse perpetually determines our selection between alternatives . . . But if [the terms on which alternatives are offered us] are altered beyond a certain point the habit will be broken or the unconscious impulse checked' (pp. 28–9). Expectations, uncertainty and consumption loans are all discussed by Wicksteed, as is the fact that rational administration of one's resources is itself costly, in terms of time and effort, and thus should not be pursued beyond a certain point. Throughout his analysis of choice between alternatives Wicksteed returns repeatedly to the idea that the most heterogeneous of satisfactions not only can be but actually are compared at the margin. He is thus led to consider how this analysis can represent 'the martyr who has borne the rack [and] is ready to be burnt to death sooner than depart a hair's breadth from the formula of his confession' (p. 404) or the man for whom there are 'certain things which he would not do for any amount of money, however large' (p. 405). Wicksteed's answer, in terms of all other considerations falling below a *minimum sensible* in such cases, appears to do little more than provide a polite reconciliation between his equality of marginal satisfactions and the presence of a *lexicographic* priority of honour over money, or of keeping the faith over escaping torture. Indeed, it is not clear how Wicksteed could maintain his own insistence that ethical considerations have priority over others (pp. 123–4), without allowing for at least some element of lexical ordering of alternatives. That said, Wicksteed's many subtle illustrations of how often disparate satisfactions *are* compared and equated at the margin remain highly instructive.

That Wicksteed pursued to the limit the concept of the rational maximizing individual is far from meaning that he had an asocial or 'atomistic' view of individual agents, or that he subscribed to the methodological fiction of the 'economic man'. On the contrary, his most important contribution to marginal theory perhaps lies in his forceful rejection of the 'economic man' concept and his closely related demonstration that the marginal analysis of individual action is entirely compatible with the recognition of the intrinsically social nature of many, even most, of the individual agent's purposes and concerns. Whilst the whole of the *Common Sense* contributes powerfully to this 'double' argument, it is in Book I, Chapter 5, 'Business and the Economic Nexus', that these issues are confronted most directly. 'But when we pass . . . to the phrase "the economic motive" . . . we are in the presence of one of the most dangerous and indeed disastrous confusions that obstruct the progress of Economics' (p. 163), Wicksteed argues, for there can be no nonarbitrary way of distinguishing motives and considerations which do influence economic actions from those which do not. There are thus two coherent alternatives; 'We may either ignore motives altogether, or may recognise all motives that are at work, according to the aspect of the matter with which we are concerned at the moment; but in no case may we pick and choose between the motives we will and

the motives we will not recognise as affecting economic conditions' (p. 165). (In fact Wicksteed very seldom adopts the former, external or behaviouristic analysis, even if there is one passage, p. 34, which strongly evokes the later 'revealed preference' approach.) If all motives are to be considered by the economic theorist, it follows, of course, that 'The proposal to exclude "benevolent" or "altruistic" motives from consideration in the study of Economics is ... wholly irrelevant and beside the mark' (p. 179); the interests which an agent seeks to pursue may or may not be directly his own. (And motivations can very well be mixed.)

But if all motives are to be taken into account, and if the principles guiding economic activity are simply the principles guiding all human activity, what defines the particular object of study of the economist? For Wicksteed, the answer lies in the concept of '*economic relations*'; 'economic investigation is concerned [with] the things a man can give to or do for another independently of any personal and individualised sympathy with him or with his motives or reasons' (pp. 4–5). When persons A and B stand in an economic relation to one another, they may well be furthering each other's purposes in fact but A enters the relation with no thought or intention of promoting B's ends and B, likewise, is motivated by no desire to further the purposes of A; however rich and complex may be the motivations of A and of B, the economic relation between them is an impersonal one. 'The economic relation does not exclude from my mind everyone but me, it potentially includes every one but you' (p. 174). To stress this point Wicksteed introduced the term 'non-tuism', which serves to focus attention upon the fact that, in an economic relation, A's lack of concern for the purposes of B (and vice versa), by no means entails that A acts from selfish motives. 'The specific characteristic of an economic relation is not its "egoism" but its "non-tuism" ' (p. 180).

With respect to the 'supply side' – a term which he might well have rejected – Wicksteed's central contributions lay in his stress on the conception of costs as opportunity costs and in his related views on reservation price and the supply curve as a 'reverse' demand curve. Wicksteed laid considerable emphasis on the idea that, no matter how indispensable productive inputs might be, 'within limits, the most apparently unlike of these factors of production can be substituted for each other at the margins' (p. 361). (Although it is noteworthy that, in the *Essay* of 1894, he had explicitly drawn attention to the possibility of completely dispensable inputs, p. 37, n. 1.) This emphasis no doubt facilitated – but did not, of course, entail – his insistence on the opportunity costs view of cost of production. 'Cost of production', he wrote, 'is simply and solely "the marginal significance of something else"' (p. 382) or, less abstractly, 'By cost of production, or cost price, when the phrase is used without qualification, I mean the estimated value, measured in gold, of all the alternatives that have been sacrificed in order to place a unit of the commodity in question upon the market' (p. 385). As he had done in 1884 and 1888, Wicksteed argued that 'there is a constant tendency to equality between price and cost of production, but not because the latter determines the former' (p. 358). The central thrust of the opportunity cost doctrine was thus directed against the 'real cost' doctrines. In his 1905 attack on the 'apologetic' school headed by Professor Marshall (referred to above), Wicksteed had written that 'To scholars of this school the admission into the science of the renovated study of consumption leaves the study of production comparatively unaffected. As a determining factor of normal prices, cost of production is co-ordinate with the schedule of demands registered on the "demand curve"'. His conclusion in 1910 was more explicit: 'The only sense, then, in which cost of production can affect the value of one thing is the sense in which it is itself the value of another thing. Thus what has been variously termed utility, ophelemity, or desiredness, is the sole and ultimate determinant of all exchange values' (p. 391). This was naturally a striking and challenging conclusion but Wicksteed did not give adequate consideration to the implications for the opportunity cost doctrine of limitations to factor mobility or of the presence of non-pecuniary benefits. (See the entry reservation price and reservation demand for further discussion of Wicksteed's 'rejection' of the supply curve.)

If Wicksteed's *Common Sense* is not flawless, it remains a brilliant demonstration that a writer who had a strongly 'social' conception of the individual agent, who was friendly to the socialist and labour movements of his time, and who was sometimes a sharp critic of the market system, could yet be a purist of marginal theory.

Selected Works

1884. *Das Kapital*: A criticism. *To-Day* NS 2, October. Reprinted in *Common Sense* (1933).

1888. *The alphabet of economic science*. Part I: *Elements of the theory of value or worth*. London: Macmillan. Reprinted, New York: Augustus Kelley, 1955.

1889. On certain passages in Jevons's theory of political economy. *Quarterly Journal of Economics*, April. Reprinted in *Common sense* (1933).

1894a. *An essay on the co-ordination of the laws of distribution*. London: Macmillan. Reprinted as No. 12 in LSE Series of Reprints of Scarce Tracts in Economic and Political Science, 1932. Reprinted, ed. I. Steedman, London: Duckworth, 1987.

1894b. Degree of utility; Dimensions of economic quantities; Final degree of utility. In *Palgrave's dictionary* of *political economy*, ed. W.-S. Jevons. Political economy and psychology. London: Macmillan. All reprinted in *Common Sense* (1933).

1905. Note on Jevons's economic work. *Economic Journal* 15, September. Reprinted in *Common Sense* (1933).

1906. Review of Pareto's *Manuale di Economia Politica. Economic Journal* 16, December. Reprinted in *Common Sense* (1933).

1910. *The common sense of political economy,* ed. L.M Robbins. 2nd edn, 2 vols. London: Macmillan. London: Routledge & Kegan Paul, 1933.

1914. The scope and method of political economy in the light of the 'marginal' theory of value and distribution. *Economic Journal* 24, March, 1–23. Reprinted in *Common Sense* (1933).

1925. Final utility. In *Palgrave's dictionary of political economy*, ed. H. Higgs, 2nd edn. London: Macmillan. Reprinted in *Common Sense* (1933).

Bibliography

Encyclopaedia of the social sciences, vol. 15, ed. E.R.A. Seligman A. Johnson. London: Macmillan, 1935.

Flux, A.W. 1894. Review of Wicksell's Über Wert, *Kapital und Rente* and of Wicksteed's essay. *Economic Journal* 4 (June): 305–313.

Herford, C.H. 1931. *Philip Henry Wicksteed: His life and work*. London: Dent & Sons.

Hutchison, T.W. 1953. *A review of economic doctrines, 1870–1929*. Oxford: Clarendon Press.

Sills, D.L., ed. 1968. *International encyclopedia of the social sciences*. Vol. 16. London: Macmillan and the Free Press.

Robbins, L. 1933. *Editorial introduction to Wicksteed* (1910).

Steedman, I. 1987. *Editorial introduction to Wicksteed* (1894a).

Stigler, G.J. 1941. *Production and distribution theories: The formative period*, 1968. New York: Agathon Press.

Sturges, R.P. 1975. *Economists' papers 1750–1950*. London: Macmillan.

Widow's Cruse

P. Bridel

The miracle of the widow's cruse (II Kings iv) was one of the most famous parables used by Keynes to present the conclusions of an analysis in the form of paradoxes.

In the *Treatise on Money*, the Fundamental Equations express the formulae for determining the price level of consumption goods and the price level of output as a whole (p. 123). The equilibrium of this two-equation system (associated with a stable price level) occurs when saving and investment balance, and is characterized by the full employment of capital and labour. To turn these 'mere identities' into cause-effect relationships. Keynes used the

traditional approach of divergences between the 'natural' and 'market' rates of interest as the factors which create 'profits' (or 'losses') by upsetting the balance between current investment and saving. On the strength of this set of assumptions, Keynes maintained that such an excess of investment over saving would be felt through an alteration in the general level of prices *only*. Hence,

> if entrepreneurs choose to spend a portion of their profits on consumption, . . . the effect is to *increase* the profit on the sale of consumption goods by an amount exactly equal to the amount of profits which have been thus expanded . . .. Thus, however much of their profits entrepreneurs spend on consumption, the increment of wealth belonging to entrepreneurs remains the same as before . . .. Thus profits, as a source of capital increment for entrepreneurs, are a widow's cruse which remains undepleted however much of them may be devoted to riotous living . . . (p. 125).

On the other hand, when saving exceeds investment and entrepreneurs make 'losses', the widow's cruse becomes a *Danaid jar* which can never be filled.

After the publication of the *Treatise*, Keynes was quick to concede that for this result to be the *only* possible effect, an assumption of *constant output* had to be added. Indeed, and partially thanks to the relaxation of this constant output assumption, Keynes managed to move away progressively from quantity adjustments grafted onto the *Treatise* argument (in which variations in the rate of interest ensure the equilibrium of planned savings and planned investment) towards changes in the level of output as *the* adjustment mechanism between saving and investment (i.e. the analytical core of the *General Theory*).

See Also

► Keynes, John Maynard (1883–1946)

Bibliography

Keynes, J.M. 1930. *A treatise on money. Vol. I: The pure theory of money.* As in *The collected writings of John Maynard Keynes*, Vol. V. London: Macmillan, 1971.

Wiener Process

A. G. Malliaris

Abstract

Brownian motion is the most renowned, and historically the first stochastic process that was thoroughly investigated. It is named after the English botanist, Robert Brown, who in 1827 observed that small particles immersed in a liquid exhibited ceaseless irregular motion. Brown himself mentions several precursors starting at the beginning with Leeuwenhoek (1632–1723). In 1905 Einstein, unaware of the existence of earlier investigations about Brownian motion, obtained a mathematical derivation of this process from the laws of physics. The theory of Brownian motion was further developed by several distinguished mathematical physicists until Norbert Wiener gave it a rigorous mathematical formulation in his 1918 dissertation and in later papers. This is why the Brownian motion is also called the Wiener process. For a brief history of the scientific developments of the process see Nelson (*Dynamical theories of Brownian motion*. Princeton: Princeton University Press, 1967).

Keywords

Bachelier, L.; Brownian motion: *see* Wiener process; Geometric Wiener process; Stochastic calculus; Uncertainty; Wiener process

JEL Classifications

C0

Having made these remarks we now define the process. A *Wiener process* or a *Brownian motion process*

$$\{Z(t,\omega) : [0,\infty] \times \mathbf{\Omega} \to R\}$$

is a stochastic process with index $t \in [0, \infty]$ on a probability space $\mathbf{\Omega}$ and mapping to the real line R, with the following properties:

(1) $Z(0, \omega) = 0$ with probability 1, that is by convention we assume that the process starts at zero.

(2) $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$ are time points then for any real set H_i $P[Z(t_i) - Z(t_{i-1}) \in H_i \text{ for } i \leq n] = \prod_{i \leq n} P[Z(t_i) - Z(t_{i-1}) \in H_i]$.

This means that the increments of the process $Z(t_i) - Z(t_{i-1})$, $i \leq n$, are independent variables.

(3) For $0 \leq s < t$ the increment $Z(t) - Z(s)$ has distribution $P[Z(t) - Z(s) \in H] = \left(1/\sqrt{2\pi(t-s)}\right) \int_H \exp\left[-x^2/2(t-s)\right] dx$.

This means that every increment $Z(t) - Z(s)$ is normally distributed with mean zero and variance $(t - s)$.

(4) For each $\omega \in \mathbf{\Omega}$, $Z(t, \omega)$ is continuous in t, for $t \geq 0$.

Note that condition (4) can be proved mathematically using the first three conditions. Here it is added because in many applications such continuity is essential. Although the sample paths of the Wiener process are continuous, we immediately state an important theorem about their differentiability properties.

Theorem (Non-differentiability of the Wiener Process)

Let $\{Z(t), t \geq 0\}$ be a Wiener process in a given probability space. Then for ω outside some set of probability 0, the sample path $Z(t, \omega)$, $t \geq 0$ is nowhere differentiable.

Intuitively, a nowhere differentiable sample path represents the motion of a particle which at no time has a velocity. Thus, although the sample paths are continuous, this theorem suggests that they are very kinky, and their derivatives exist nowhere. The mathematical theory of the Wiener process is presented rigorously in Billingsley (1999, ch. 37) and more extensively in Knight (1981).

The first application of Brownian motion or the Wiener process in economics was made by Louis Bachelier in his dissertation 'Théorie de la spéculation' in 1900. Cootner (1964) collects several papers and cites additional references on the application of the Wiener process in describing the random character of the stock market. In the early 1970s Merton, in a series of papers, established the use of stochastic calculus as a tool in financial economics. The Wiener process is a basic concept in stochastic calculus and its applicability in economics arises from the fact that the Wiener process can be regarded as the limit of a continuous time random *walk* as step sizes become infinitesimally small. In other words, the Wiener process can be used as the cornerstone in modelling *economic uncertainty* in continuous time. For purposes of illustration consider the stochastic differential equation

$$dX(t) = \mu(t, x)\,\mathrm{d}t + \sigma(t, x)\,\mathrm{d}Z(t) \qquad (1)$$

which appears in the economic literature describing asset prices, rate of inflation, quantity of money or other variables. In (1), changes in the variable $X(t)$, denoted as $dX(t)$, are described as a sum of two terms: $\mu(t, x)$ which is the expected instantaneous change and $\sigma(t, x)\, dZ(t)$which is the unexpected change. Furthermore, this unexpected change is the product of the instantaneous standard deviation $\sigma(t, x)$ and uncertainty modelled by increments in the Wiener process. See Merton (1990, ch. 3) for a methodological essay on continuous-time modelling, and Malliaris and Brock (1982) or

Chang (2004, ch. 2) for numerous applications of the Wiener process in economics and finance.

Economists have constructed various processes based on the Wiener process. Let $\{Z(t),\ t \geq 0\}$ be a Wiener process and use it to construct a process $\{W(t),\ t \geq 0\}$ defined by $W(t) = Z(t) + \mu t,\ t \geq 0$ where μ is a constant. Then we say that $\{W(t),\ t \geq 0\}$ is a *Wiener process* or *Brownian motion process with drift* and μ is called the drift parameter. In this case the only modification that occurs in the definition of a Wiener process is in property (3) where $W(t) - W(s)$ is normally distributed with mean $\mu(t - s)$ and variance $(t - s)$. Finally, let $W(t)$ be a Wiener process with drift as just defined. Consider the new process given by $Y(t) = \exp[W(t)],\ t \geq 0$. Then $\{Y(t),\ t \geq 0\}$ is called a *geometric Brownian motion* or *geometric Wiener process*.

The availability of an extensive mathematical literature on the Wiener process and the economists' fundamental goal to model economic uncertainty in continuous time suggest that this process will continue to be an important tool for economic theorists.

See Also

- ▶ Martingales
- ▶ Uncertainty

Bibliography

Billingsley, P. 1999. *Probability and measure*. 3rd ed. New York: Wiley.

Chang, F.R. 2004. *Stochastic optimization in continuous time*. New York: Cambridge University Press.

Cootner, P.H. 1964. *The random character of stock market prices*. Cambridge, MA: MIT Press.

Knight, F.B. 1981. *Essentials of Brownian motion and diffusion*, Mathematical surveys, number 18. Providence: American Mathematical Society.

Malliaris, A.G., and W.A. Brock. 1982. *Stochastic methods in economics and finance*. Amsterdam: North-Holland Publishing Company.

Merton, R.C. 1990. *Continuous-time finance*. Malden: Blackwell.

Nelson, E. 1967. *Dynamical theories of Brownian motion*. Princeton: Princeton University Press.

Wieser, Friedrich Freiherr, (Baron) von (1851–1926)

E. Streissler

Keywords

Austrian economics; Austrian School; Böhm-Bawerk, E. von; Countervailing power; Efficient production; Equimarginal principle in production; Imputation; Innovation; Marginal utility; Menger, C.; Mises, L. E. von; Natural value; Opportunity cost; Schumpeter, J. A.; Wieser, F. F. von

JEL Classifications

B31

Wieser is commonly cited together with his senior, Carl Menger, and his exact contemporary, Eugen Böhm von Bawerk, as one of the founding trio of the Austrian School of Economics in the last quarter of the 19th century. The exact nature of his achievement, however, seems now practically forgotten: possibly because he produced an intractable mixture of deep and influential insights, very distinctly his own, intermingled with oratorical prose and often unpalatable value judgements; he was extremely successful in his own generation but appeared outdated in his attitudes half a century later.

Wieser was born on 10 July 1851, in Vienna. His father was Commissary-General of the Austrian army in the war of 1859, for which service he was ennobled, later becoming Vice President of the Austrian Court of Audit, a baron and a privy councillor (Geheimrat). But this high social status was only acquired after Friedrich Wieser's birth and very little money went with it so that the family lived in modest circumstances. Wieser went to the Benedictine Schottengymnasium in Vienna, one of the city's three elite schools. His classmate was Eugen Böhm von Bawerk, who became his close

friend and brother-in-law. Together the two studied at Vienna University law faculty (which included courses in economics), together they entered the civil service in the fiscal division, and together they went on a two-year leave of absence to perfect themselves in economics at Heidelberg, Leipzig and Jena, with Knies, Roscher and Hildebrand. A little after Böhm, Wieser passed his 'Habilitation' in economics with Menger in 1883, was appointed associate professor in 1884 and full professor in 1889 at the University of Prague and was that university's Vice Chancellor in 1901–2. In 1903, he succeeded Menger in the chair of economic theory at Vienna University law faculty on the latter's early retirement, Böhm again joining him only a year later as extraordinarily appointed additional full professor. Böhm, Menger, and finally also Wieser served as members of the Austrian House of Lords (Herrenhaus). Wieser became Minister of Commerce in 1917, holding this office up to the end of the monarchy in 1918. He died on 23 July 1926.

Apart from the short ministerial interlude Wieser thus taught from 1884 to 1926 at the largest universities of Austria. Basically, he must be considered the most successful teacher (especially of undergraduates) and the orator of the trio. His influence was pervasive through his lecturing to tens of thousands of law students, many of whom he examined in person, and at second and third remove on even vaster numbers in the intellectual melting pot of Vienna. In a true oral tradition Wieser influenced present-day economics through what these frequently very young students – an appreciable percentage of whom later became themselves important in Western intellectual life – picked up in a kind of intellectual osmosis, mostly without realizing it and therefore usually without attributing the ideas to their teacher.

Wieser's main works are: his thesis of 'Habilitation' (Wieser 1884), which encompasses a large part of his original thought, particularly the marginal productivity valuation of factors of production and his cost theory; Wieser (1889), mainly an elaboration of the former together with an attempt to give the marginal utility concept normative distributional content; Wieser (1914), the definitive textbook of the Austrian School and (with its rival, G. Cassel's more up-to-date *Theoretische Sozialökonomie,* 1918) one of the two main theoretical textbooks in German of the early interwar period, a book still worth reading, especially the less well-known institutional chapters on large corporations and money and banking; finally Wieser (1926), a socio-philosophical tract in abject adulation of power (power being justified by mere 'success'), whose lack of judgement can only be justified by the effect of the total breakdown of all established social and political order after the First World War on Wieser's own moral fibre.

Wieser prided himself on the invention of telling phrases, particularly the term 'Grenznutzen' in Wieser (1884), wherefore Marshall credits him, perhaps unjustly, with originating the term 'marginal utility'. During his leadership the Austrian School had to sail under the flag 'Grenznutzenschule'. In contrast to the purely analytic usage by the other members of the trio, 'Grenznutzen' had for Wieser a near mystic connotation and certainly normative content: more precisely, it is the average marginal utility in a competitive society with equality of incomes which is the 'natural value' of Wieser (1889). 'Grenznutzen' thus served Wieser, who was (unusual for an Austrian economist) a paternalistic interventionist, as a yardstick of policy evaluation.

In contrast to Menger and Böhm, Wieser was not a clear logical analyst but had influential visions. He was clearest in his cost and production theory, frequently being credited with introducing the opportunity cost principle that all costs are only utilities forgone, though Wieser's actual advance over Menger appears slight. Wieser certainly, however, gave what appears to be the first account of the principles of efficient production, which Menger had ignored (Wieser 1884). Production is undertaken in expectation of the price the marginal valuation of consumers will allow, Wieser (1884) first formulating the equimarginal principle in production: the marginal product of each factor (or its cost) must be the same in all its different uses and as high as the least important marginal utility achievable from its given supply (Wieser's Law of Cost). In Wieser (1914) this is extended (contemporaneously with Wicksteed, *Common Sense of Political Economy,* 1910) to

an analysis of differential quality rents on the lines of Ricardo: any more efficient factor earns as rent the additional value added over the least efficient equivalent factor. Some of his insights into capital and efficient production Wieser probably owes, as the terminology suggests, to Marx (Wieser never gives his authorities, apart from sparse references in Wieser, 1914): for example that the value of factors of production must reflect the socially necessary cost of production (the use of the best generally known technique); or that innovation brings extra profits to the innovator, without changing the value of the (other) factors. (Marx, Engels, Ricardo, Jevons and Menger are the five authors Wieser acknowledges as inspiration in the Introduction to Wieser, 1884.) As to distribution, Wieser first posed the problem whether the marginal product reward of all factors of production would exactly exhaust the product ('Zurechnungsproblem', the problem of imputation), without being, despite many attempts, able to solve it.

These ideas were, however, all on the point of being discovered by others. Uniquely his own is the repeated stress, already in Wieser (1884), of the paramount importance of economic calculation and the need to have an economic measuring rod for all rational 'planning' for the future. (One is tempted to suspect this to be an obsession of the son of the Vice President of the Court of Audit.) The measuring rod for Wieser is marginal utility in its wide sense; but it was a small step, taken by Mises and Hayek, to make out of this need for a measuring rod in all economic planning the concept of the informative nature of prices. Economics may even owe the (then uncommon) term 'planning' for the rational activity in economics on the individual as well as the societal level to Wieser via Mises. Wieser already stated repeatedly that even a socialist economy would have to use the same economic measuring rod and basically the same principles of 'planning' as a capitalist one: out of which Mises developed the idea that, lacking prices, a socialist society could not plan rationally.

Besides his production and distribution-oriented ideas Wieser had a second influential vision: the importance of the creative individual in all economic processes. He felt deeply the basically contradictory nature of his two visions, the impersonal mass effects of efficient production on the one hand, an idea which he curiously traced to the influence of Herbert Spencer, and the elitist idea of the effects of the outstanding individual, which he attributed to the hero-worshipping teaching of history in the Schottengymnasium. In this vein, which he cultivated in his later years, he was again, above all, influential through the forceful and suggestive use of words, the terms 'Führer', 'Pionier', 'Neuerung' (German for innovation) being of his creation. Schumpeter adopted virtually all the terminology for his *Theory of Economic Development* (1912) from his acknowledged teacher Wieser and also the idea that economic 'dynamics' (in contrast to statics) is due to individual leadership activity. Wieser himself had developed relatively few concrete conclusions out of his leadership rhetoric, apart from remarks about the countervailing power of trade unions and the administrative and even innovative efficiency of large corporations in Wieser (1914), an idea taken up by Schumpeter only later. Wieser's second vision degenerated into the lurid prose of Wieser (1926), where, for example, the 'Führer' (a pet word of Wieser's), Adolf Hitler, is chided (in 1926!) for not quite making the grade. For Wieser, again in sharp contrast to the staunch liberal principles of Menger and Böhm, tended, in spite of his basic Catholic-conservative outlook, to flirt with any social movement that was new and appeared 'great', making commendatory references to socialism in his youth and to German nationalism and fascism in his old age.

See Also

► Austrian Economics

Selected Works

1884. *Über den Ursprung und die Hauptgestze des wirtschaftlichen Werthes.* Vienna: Hölder.
1889. *Der Natürliche Werth.* Vienna: Hölder. Trans. as *Natural value,* ed. W. Smart. London: Macmillan, 1893. Reprinted, New York: Kelley & Millman, 1956.

1914. Theorie der gesellschaftlichen Wirtschaft. In *Grundriss der Sozialökonomik,* vols. 1 and 2. Tübingen: Mohr-Siebeck. 2nd ed., 1923. Trans. as *Social economics.* New York: Greenberg, 1927.
1926. *Das Gesetz der Macht.* Vienna: Springer.

Bibliography

Mayer, H. 1929. Friedrich Freiherr von Wieser. In *Neue österreichische Biographie*, vol. 6. Vienna: Amalthea.
Schumpeter, J.A. 1912. *The theory of economic development. An inquiry into profits, capital, credit, interest, and the business cycle.* Trans. R. Opie. Cambridge, MA: Harvard University Press, 1934.
Stigler, G.J. 1941. *Production and distribution theories: The formative period*. New York: Macmillan.
Streissler, E.W. 1986. Arma virumque cano – Friedrich von Wieser, the bard as economist. In *Die Wiener Schule der Nationalökonomie*, ed. N. Leser. Vienna: Hermann Böhlau.

Williams, John Henry (1887–1980)

Stephen V. O. Clarke

John H. Williams was a member of the US delegation to the Preparatory Committee for the World Monetary and Economic Conference of 1932–3; Nathaniel Ropes Professor of Political Economy, 1933–57 and Dean of the Graduate School of Public Administration, Harvard University, 1937–47; Vice President and, subsequently, Economic Adviser, Federal Reserve Bank of New York, (1936–56) and President of the American Economic Association in 1951.

Williams was a rare combination of the scholar and the practitioner of the art of central banking. Familiar with the evolution of economics, his career was devoted to the application of that discipline to public policy.

In doing so, he saw value, as well as danger, in theory. Without theory, as Keynes had said, we are 'lost in the woods'. But theory inevitably simplifies reality. Moreover, the most influential theories originate in unique circumstances and in views about policies to deal with them. Since circumstances change, he warned that those who prescribe glibly from theory are dangerous as policy-makers.

His views about theory were applied with the greatest effect in the international monetary sphere. He particularly questioned conventional views about the gold standard. The classical specie flow mechanism was a beautiful intellectual construct. However, the pre-1914 reality was that Britain maintained a gold standard while related countries based their currencies on sterling. With the subsequent rise of rival centres, the maintenance of international monetary stability boiled down to negotiating mutually acceptable relations between the 'key currencies', and then maintaining those relations – stable but not immutable – through appropriate domestic policies in the centre countries. Such views clearly influenced the negotiation of the Tripartite Agreement of 1936. They were also the basis for Williams' reservations, a few years later, about the International Monetary Fund.

A selection of Williams's more important works is published in his *Postwar Monetary Plans and Other Essays* (1945).

Selected Works

1945. *Postwar monetary plans and other essays*. New York: Knopf.

Williamson, Oliver E. (Born 1932)

Scott E. Masten

Abstract

Oliver E. Williamson is the 2009 co-recipient (with Elinor Ostrom) of the Nobel Memorial Prize in Economics, awarded 'for his analysis of economic governance, especially the boundaries of the firm'.

Keywords

Adaptation; Antitrust; Asset specificity; Barriers to entry; Bounded rationality; Contract law; Corporate finance; Corporate governance; Credible commitments; Firm boundaries; Forbearance law; Franchise contracting; Fundamental transformation; Governance; Incomplete contracting; Institutions; Managerial discretion; Organisation; Opportunism; Predatory pricing; Regulation; Remediableness; Selective intervention; Transaction cost economics; Vertical integration

JEL Classifications

B31

Introduction

Oliver E. Williamson was born in 1932 in Superior, Wisconsin. He received a BSc in engineering from the Massachusetts Institute of Technology in 1955, an MBA from Stanford University in 1960, and his PhD in economics from Carnegie-Mellon University in 1963. Williamson began his academic career at the University of California, Berkeley, moving to the University of Pennsylvania in 1965 and to Yale University in 1983. In 1988, he returned again to Berkeley, where he is currently Edgar F. Kaiser Professor Emeritus of Business, Economics, and Law.

Early Career and Contributions

Entering the doctoral program at Carnegie in 1960 as a transfer from the PhD program in business administration at Stanford, Williamson was attracted to the issues being addressed by the behavioural group led by Herbert Simon, Richard Cyert and James March but favoured the methods employed by the economists, which included John Muth, Merton Miller and Allan Meltzer. His dissertation combined those interests to analyse the behaviour of firms where managers had preferences over such things as staff and emoluments and the cost of detecting and policing managers' actions prevented owners from effectively overseeing managers. This research, awarded a Ford Foundation Dissertation Prize in 1963 and published as *The Economics of Discretionary Behavior* (1964), introduced the first consistent and economically sound model of firm behaviour based explicitly on utility maximisation rather than profit maximisation and opened for the first time the black box of the firm to the tools of modern economics.

Williamson went on to make a number of important contributions to economics and public policy, including original and influential papers on peak-load pricing (1966), social choice (1967b), and the dynamics of inter-firm behaviour (1965), as well as a pair of articles on barriers to entry, 'Selling expense as a barrier to entry' (1963) and 'Wage rates as a barrier to entry: the Pennington case in perspective' (1968), that were the first to treat entry barriers as a strategic decision and were forerunners to the later strategic barriers literature. In 'Economies as an antitrust defense' (1968) Williamson demonstrated that potential cost economies from horizontal mergers could easily outweigh the dead-weight losses from increased market power and advocated that antitrust policy be modified to recognise demonstrable cost economies as a valid defence in merger cases. Recognition of the potential for antitrust policy to impede efficient behaviour was also evident in his efforts to develop a practical rule for distinguishing competitive from predatory behaviour (1977). The resulting quantity-based rule both exhibited superior

welfare properties and was easier to implement than previously proposed rules based on the comparison of prices and costs.

Williamson's primary interest, however, remained issues of organisation, and he continued to develop and refine his analysis of the relation between the internal structure and behaviour of firms. Intrigued by Alfred Chandler's description of the advent of the multidivisional corporation, Williamson set out to analyse how growth in the size and complexity of a firm affected the ways in which tasks and responsibilities are divided within the organisation (1967a, 1970). Comparing unitary, or U-form, organisation, in which operations are grouped along functional lines (sales, finance, manufacturing, and so forth), with multidivisional, or M-form, organisation, in which decision-making responsibility was assigned to quasiautonomous operating divisions organised along product, brand or geographic lines, Williamson argued that the strains on management inherent in large organisations favoured M-form organisation, which, appropriately administered, had the properties of a miniature internal capital market. The analysis resulted in the M-form hypothesis: 'the organization and operation of the large enterprise along the lines of the M-form favors goal pursuit and least-cost behavior more nearly associated with the neoclassical profit maximization hypothesis than does the U-form organizational alternative' (1970, p. 134).

Transaction-Cost Economics

Williamson's crowning achievement, however, for which he received the Nobel Prize, was his development of the economics of transaction costs. At the heart of the transaction cost approach is the insight, originally discerned by Ronald Coase (1960), that, in a world of zero transaction costs, individuals would always negotiate to realise all potential gains from trade or cooperation regardless of organisational form or the prevailing institutions. Governance form matters, therefore, only to the extent that (i) transaction costs prevent individuals from reaching and enforcing mutually advantageous bargains and (ii) the nature or size of those costs differ between institutions or organisational forms. Although Coase's insight provided the logical foundation for focusing on transaction costs, the difficulty of observing and measuring transaction costs in practice, and the fact that the costs of transacting can, at best, only be observed for governance arrangements actually chosen, subjected early transaction cost arguments to the criticism that efficiency claims based on transaction costs were easy to make and impossible to refute (see, e.g. Fischer 1977; Simon 1991).

The key to 'operationalising' the concept of transaction cost, Williamson reasoned, was to (i) identify the critical features distinguishing governance alternatives, (ii) assess the differential capacities and limitations of those alternatives in relation to attributes of transactions, and finally, (iii) match governance structures and transactions 'in a discriminating way', thereby allowing predictions relating (observable) transaction attributes and organisational form to be formulated and tested (Williamson 1979, pp. 234, 261). Following that prescription required substantial departures from the established framework of neoclassical economics, which assumed a strong form of rationality and costless contract enforcement. In their place, Williamson substituted the concepts of *bounded rationality* and *opportunism*. Bounded rationality (a term introduced by Herbert Simon 1957, 1961), refers to the fact that, although individuals intend to act in a rational manner, their ability to realise those intentions is hampered by their limited knowledge, foresight and computational ability. Because of bounded rationality, individuals cannot solve complex problems instantaneously, anticipate all possible future events, or reliably devise and communicate appropriate responses to contingencies they do foresee. Moreover, because everyone suffers such limitations, there exist no omniscient third parties capable of optimally planning and directing economic activity (hierarchy) or of resolving disagreements between transactors accurately and cheaply (courts). Second, to account for the possibility of strategic defection from agreements, Williamson introduced the concept of opportunism: the willingness of (at least

some) transactors to renege on promises, cheat on agreements, shirk responsibilities, circumvent rules, search out loopholes, or otherwise exploit the vulnerabilities of another in the hope of obtaining a larger share of benefits of the exchange. Because efforts to control opportunism invariably place additional demands on bounded rationality, the goal of economic organisation becomes to 'organize transactions so as to economize on bounded rationality while simultaneously safeguarding them against the hazards of opportunism' (1985, p. 32).

Having identified the human factors responsible for organisational failures, it remained to show how the incidence of those failures related to features of the economic environment. Central to this task was Williamson's recognition that even bounded rationality and opportunism pose few organisational problems in a static world. 'Transactions conducted under certainty are relatively uninteresting. Except as they differ in the time required to reach an equilibrium exchange configuration, any governance structure will do' (1979, pp. 253–4). In the presence of change and uncertainty, however, transactors need to plan, monitor and continually adjust their behaviour, activities that demand attention and often cooperation. Thus, for Williamson, change is the pivotal feature of the economic environment and adaptation to unfolding events the central problem of economic organisation (1975, p. 5, 1991, pp. 277–8).

Economic institutions represent alternative ways of governing the process of adaptation and differ both in their capacities to effect adaptations and in the costs associated with doing so. Discrete market transactions, for instance, provide transactors considerable autonomy and flexibility in the periods both leading up to and following the actual transaction (1979, 1991). In such transactions, parties are generally free to bargain or not bargain as they please and, once the transaction is consummated, have relatively few ongoing obligations. The latitude afforded transactors in simple market transactions provides them both the incentive and the ability to adjust their behaviour to unfolding events. But it also furnishes a variety of tactics through which transactors may seek to elicit a more favourable distribution of the gains from trade. Parties to a simple exchange may haggle, stall or walk away from a deal altogether in hopes of extracting more of the rents accruing to exchange.

Such opportunistic tendencies matter little where the identity of traders is unimportant; the scope for opportunism is limited where transactors can easily turn to alternative trading partners if one seeks to gain at the expense of the other. But realisation of cost economies or design benefits often requires investments in relationship-specific assets that isolate the transactors from market alternatives. Relationship-specific investments can take at least four forms: (i) physical asset specificity, which involves investments in equipment such as tooling or dies specially designed to serve a particular customer; (ii) site or location specificity, which occurs when a buyer or seller locates his or her facilities next to the other to economise on transportation costs; (iii) human asset specificity, which arises when one or both parties develop skills or knowledge valuable only when dealing with the other; and (iv) dedicated assets, which are investments made to support exchange with a particular customer that, though not specific to that customer, would result in substantial excess capacity were the customer to discontinue purchases (Williamson 1983, p. 526). Once the die is cast and physical or human capital has been specially designed or located for a particular use or user, continuity in trading relationships becomes important. This 'Fundamental Transformation' from a situation of *ex ante* competition to small numbers bargaining when relationship-specific investments are made becomes a dominant force motivating the adoption of specialised governance structures in Williamson's framework (Williamson 1985, pp. 61–3): Without some form of safeguard against appropriation, parties will be reluctant to invest in relationship-specific assets, despite the gains from doing so, for fear that those gains will be dissipated in subsequent contention over their distribution.

Securing the terms of trade at the outset through a long-term contract is one such safeguard. But contracting increases the demands on bounded rationality and only imperfectly limits

opportunism. To accommodate uncertainty, contractors must either anticipate and devise responses to a large number of contingencies or prescribe a process through which adaptations can be executed. They must do so, moreover, in terms that courts can be expected to understand and implement at reasonable cost. The difficulty of anticipating and defining contractual obligations that avoid the prospect of costly adjudication means that contracts will, on the one hand, tend to be inflexible and, on the other, leave considerable opportunity to cheat on the agreement or otherwise seek to evade performance (1983, pp. 526–7, 1985, p. 21). As transactions become more complex and the environment more uncertain, the limitations of contracting as a safeguard against opportunism grow, increasing the attraction of other institutional arrangements that better support adaptive, sequential decision making while circumscribing or redirecting opportunistic tendencies.

The Theory of the Firm

Williamson first conceived and applied this framework for analysing economic organisation in the context of the problem that had also originally led Coase (1937) to focus on transaction costs, namely, the nature and boundaries of the firm. The orthodox portrayal of the firm as a production function that combines inputs purchased from individuals and other firms into outputs that it in turn sells on the market had proved useful for the analysis of market equilibrium. But this abstraction failed to illuminate the purpose or consequences of the vast array of economic activity organised administratively within firms. In the absence of an efficiency rationale, integration of production tended to be seen either as technologically determined or as serving monopoly purposes.

Transaction-cost economics recast the firm as a governance structure, one among several alternative ways in which production and exchange might be organised. By integrating a transaction, transactors alter the rules and processes through which disputes are resolved and adjustments effected. But though a firms-as-governance-structure orientation was a significant step, a complete theory of the firm awaited resolution of three great puzzles. First, what are the properties that distinguish organisation within the firm from market exchange? Second, what determines which transactions get integrated? And third, what limits firm size or, as Coase (1991 (1937), p. 23) put it, 'why is not all production carried out in one big firm?'.

Williamson's initial assault on the boundaries of the firm issue, 'The vertical integration of production: market failure considerations' (1971), introduced the essential elements later to become cornerstones of the comparative institutional framework. In particular, Williamson (i) described the advantages of markets in providing incentives and reducing the demands on boundedly rational decision makers; (ii) explained why long-term, complete contingent-claims contracts were not feasible; and (iii) identified the hazards of short-term contracting 'if either (1) efficient supply requires investment in special-purpose, long-life equipment, or (2) the winner of the original contract acquires a cost advantage, say by reason of "first mover" advantages (such as unique location or learning, including the acquisition of undisclosed or proprietary technical and managerial procedures and task-specific labor skills)' (1971, p. 116). 'In circumstances. where protracted bargaining between independent parties to a transaction can reasonably be anticipated', internalisation offered 'a wider variety and sensitivity of control instruments' that included 'lowcost access to the requisite data' and 'a comparatively efficient conflict resolution machinery' (1971, p. 113).

It was not until the publication of *Markets and Hierarchies* in 1975, however, that the importance and scope of Williamson's analysis of organisation began to be fully appreciated. In this path-breaking book, Williamson organised the components of his earlier analysis into a unified and systematic framework for analysing the problem of economic organisation in comparative terms. Among other things, he categorised the behavioural and environmental attributes responsible for organisational failures, described the

distinctive features of market and hierarchical organisation, and analysed the differential effects of idiosyncratic investments and uncertainty on the costs of governing intermediate product transactions internally versus externally. The problems posed by incomplete or 'impacted' information – which Williamson traced to their rudiments in bounded rationality and complexity – also played a central role (1975, pp. 31–7). The publication of *Markets and Hierarchies* was a landmark in the development of transaction-cost reasoning that opened the door to the investigation of a host of organisational problems that had previously resisted economic analysis.

Williamson continued the refinement and extension of transaction-cost reasoning as well as his analysis of the governance of intermediate product transactions in numerous articles and books, including *The Economic Institutions of Capitalism* (1985) and *The Mechanisms of Governance* (1996). Of particular importance in this progression was his 1979 article, 'Transaction-cost economics: the governance of contractual relations' (1979; at the time of this writing, the third most cited article in the *Journal of Law and Economics*, behind only Coase's 'The problem of social cost' (1960) and one other). In addition to offering a refined statement of the logic of matching governance arrangements with transactions, this article identified the essential relation between governance modes and contract law regimes.

Through his efforts to operationalise transaction-cost theory, Williamson had provided by this point an answer to the question: what determines which transactions get integrated? He had, moreover, as part of that analysis, also described some of the properties that distinguish internal organisation from markets (1975, pp. 29–30) and the limits to internal organisation (1975, ch. 7). But the basis for the superior auditing and dispute resolution properties that Williamson and others ascribed to internal organisation remained a matter of controversy. Where does the authority of management to direct production or settle disputes come from, and why are employees less able to hide or distort information than independent contractors? As Alchian and Demsetz had earlier protested, employers have no authority or disciplining power beyond that available in any ordinary contractual relationship; all an employer can do is 'fire or sue' (1972, p. 777). A fully satisfactory accounting of the source of internal organisation's distinctive properties and the limits to firm size remained elusive.

Williamson supplied the last two pieces of the integration puzzle with the notions of 'forbearance law' and 'the impossibility of selective intervention'. Building on his earlier association of governance modes and contract regimes (1979), Williamson proposed that the distinctive feature of the law governing internal organisation is forbearance: 'whereas courts routinely grant standing to firms should there be disputes over prices, the damages to be ascribed to delays, failures of quality, and the like, courts will refuse to hear disputes between one internal division and another over identical technical issues' (1991, p. 274). The decision to make rather than buy thus has substantive implications: whereas parties to a contract can resort to courts to resolve disputes, top management exercises ultimate authority in disputes between divisions; it is its own court of last resort. Ultimately, termination and legal action are, as Alchian and Demsetz maintained, the only options available, but when you can fire and what you can sue for depend on the mode of organisation adopted. The refusal of courts to intervene in internal disputes affords management flexibility to conduct business and adapt operations as they deem appropriate and thus provides a basis for the control and adaptability advantages of internal organisation.

But if internal organisation possesses superior control and adaptation properties, why are not all transactions organised with the firm? In principle, a newly integrated firm should be able to operate at least as efficiently as the two independent firms from which it was formed simply by allowing each division of the combined firm to operate independently as it had before and only intervening where net benefits were likely to be realised. As long as managers intervened selectively, combined operations would always dominate independent ones (1985, pp. 131–5). Williamson's answer to the question, 'what limits firm size?',

lay in the impossibility of selective intervention (1985, ch. 6). Unable to use the courts to enforce promises to intervene selectively, management would be drawn to intervening even where joint benefits are not realised. Without effective assurances that owners will not appropriate performance enhancements, the incentives of division managers to innovate, maintain assets, acquire and utilise information, and otherwise invest in the efficient operation of the division are ineluctably compromised. In their place, the firm is forced to substitute weaker, indirect incentives dependent on managerial oversight. The loss of incentive intensity combined with the limited capacity of management to administer additional transactions – which manifest themselves in a variety of bureaucratic inefficiencies – ultimately undermine the efficacy of internal organisation and thereby limit firm size.

With the last pieces of the puzzle in place, the tradeoffs between market and internal organisation came into still sharper focus. Although the high-powered incentives attainable with market exchange economise on bounded rationality, the dissociation of effort and compensation and resulting loss in incentive intensity resulting from integration is not always to be lamented: high-powered incentives motivate efforts to redistribute as well as increase rents. Where asset specificity is great and uncertainty high – hence the gains to ongoing exchange large and flexibility highly valued – flatter, low-powered incentives supported by enhanced monitoring are likely to be preferred (1988a, 1991).

Applications, Extensions, and Recurring Themes

Williamson recognised early on that the transaction cost framework, although developed in the context of the theory of the firm, was far more general, maintaining that 'any issue that can be formulated as a contracting problem can be investigated to advantage in transaction cost terms' (1985, p. 17), which is to say, wherever gains from trade or cooperation arise. Except in cases where all costs are incurred and benefits accrue simultaneously, transactors who move first are vulnerable to reneging by those who move later. Without adequate assurances that others will, when the time comes, uphold their end of the bargain, public utilities will be reluctant to invest in infrastructure, benefactors may withhold donations, legislators may fail to strike deals to enact legislation, and citizens may forgo wealth creation for fear, in each case, that the other side will appropriate the value of their investment. In these, as in commercial transactions, the benefits of trade and cooperation will often be enhanced if the parties can find ways to make their commitments credible even if, as is often the case, courts are ineffective or simply unavailable. In 'Credible commitments: using hostages to support exchange' (1983), Williamson noted that, even when courts are available, the cost and imperfections inherent in court enforcement will often lead contracting parties to seek out devices that foster cooperative adaptation to change without the need for recourse to the court system. The use of economic hostages is one such device: as in the days when kings extended their daughters as collateral against breach of their commitments to other monarchs, a modern commercial transactor might find it advantageous to make relationship-specific investments whose value would be sacrificed if he failed to perform as promised. Although a unilateral investment in relationship-specific assets exposes the transaction to appropriation hazards, a reciprocal investment by a trading partner that balances the parties' exposure to such hazards may strengthen the integrity of a trading relationship, suggesting efficiency motivations for a range of otherwise enigmatic practices, such as reciprocal dealing and aspects of franchise contracts. Similarly, political, regulatory, and economic institutions can often be understood to promote cooperation and trade by enhancing the credibility of commitments among the corresponding parties.

Another recurring theme in transaction cost analyses is the need for comparative analysis. Whereas conventional analyses assess the efficiency of organisational arrangements relative to the absolute standard of Pareto optimality, transaction-cost economics maintains that the

merits of particular organisational arrangements can *only* be assessed relative to the performance of the relevant alternatives constrained by the same human frailties and propensities (bounded rationality and opportunism), technology, and information (Williamson 1985, 1993a). A transaction-cost orientation thus fosters an appreciation that problems ascribed to markets, contracts or regulation often inhere in the circumstances, not in the institutions. Moreover, given the powerful motives of individuals to find and adopt governance arrangements that increase the available surplus, enduring organisational forms and institutions deserve the (rebuttable) presumption of efficiency, a notion that Williamson has codified as the remediableness criterion: 'An extant mode of organization for which no superior *feasible* form of organization can be described and *implemented* with expected net gains is *presumed* to be efficient' (Williamson 1996).

Williamson's framework has been applied extensively in, among other fields, law and political science as well as economics. Transaction-cost economics has also become one of the dominant paradigms in strategic management research: by providing a systematic way of analysing the relative merits of alternative governance arrangements and a set of testable propositions relating those merits to attributes of transactions and the surrounding environment, transaction-cost economics offers strategy a set of normative rules for choosing among alternative organisational arrangements. An empirical literature testing transaction-cost hypotheses, which now contains hundreds of studies, has been broadly corroborative (e.g. Macher and Richman 2008).

Among Williamson's own applications and extensions of his framework are analyses of franchise contracting versus regulation for public utilities and of corporate governance. In a challenge to the conventional wisdom of the late 1960s – that public utility regulation was a necessary response to natural monopoly cost conditions – Harold Demsetz (1968) proposed that public utility services could be efficiently procured (even where cost conditions dictated a single supplier) simply by awarding a franchise to the firm that offered to serve the market at lowest price. Williamson, perceiving parallels to the theory of vertical integration on which he had been working, responded that, because the supply of public utility services typically requires large, durable investments in production and distribution facilities that are specialised to a particular market, efficient franchise agreements would have to take the form of long-term contracts to avoid repeated haggling over the terms of trade once those investments were in place. But uncertainty about cost and demand conditions over such long horizons and the complexity of public utility services would leave long-term contracts for public utility services perilously incomplete. To accommodate that uncertainty, franchise contracts would have to employ contract terms and administrative machinery – cost-plus price adjustment, auditing procedures, elaborate and formal dispute resolution processes – that mirrored both in character and in costs the administrative apparatus traditionally associated with rate-of-return regulation (Williamson 1976). Whereas Demsetz had revealed the potential for the efficient supply of public utility services using market arrangements if 'irrelevant complications' were ignored (Demsetz 1968, p. 57), Williamson had shown that it was precisely such complications that underlay the choice between regulation and franchise bidding and that the complications that impeded effective regulation were likely to frustrate franchise bidding solutions as well.

Williamson also showed how a transaction-cost orientation could be used to provide a fresh perspective on firms' financing decisions (1988b). Debt and equity, he argued, can viewed as alternative governance structures for the procurement of financial capital, analogous to contracting and internal organisation for the procurement of intermediate products. For projects involving standardised redeployable assets, debt, which offers contractual protections for investors, is the low-cost governance arrangement. As asset specificity increases, however, the residual value of the assets and, hence, the value of debt holders' preemptive claims, declines. The willingness to supply capital in such cases will be enhanced if management offers investors a safeguard against appropriation or misuse of their investments in the

form of a body (the board of directors) that (i) can monitor the use of investors' capital and (ii) has the authority to oversee, compensate and replace management. Like internal procurement of intermediate goods, and unlike debt, which affords investors the power to intervene only under a set of relatively extreme events (such as bankruptcy), equity financing provides a mechanism for regularised interventions. In contrast to traditional agency treatments, transaction cost economics 'regards debt and equity principally as governance structures rather than as financial instruments' (1988b, p. 579).

Conclusion

Williamson has consistently advocated viewing organisation through the 'lens of contract' instead of through the orthodox 'lens of choice' while maintaining that understanding governance is an interdisciplinary project that combines economics with organisation theory and law. His career is a testament to the productivity of such an approach. By relating the advantages and liabilities of alternative organisational arrangements to features of the transaction in a discriminating way, Williamson demonstrated that a transaction-cost orientation could indeed generate refutable implications. In doing so, moreover, he provided a systematic conception of the problem of economic organisation in which all organisational forms are recognised to be subject to the same fundamental limitations but, at the same time, permits the assignment of 'transactions (which differ in their attributes) to governance structures (the adaptive capacities and associated costs of which differ) in a discriminating (mainly transaction cost economizing) way' (1989, p. 136). The result has been to alter irreversibly the way we perceive and analyse organisation.

Further Reading

For more on the development and contributions of Williamson's scholarship, see Williamson (1995, 2010) and Masten (1996).

See Also

- ▶ Barriers to Entry
- ▶ Firm, Theory of the
- ▶ Transaction Costs, History of
- ▶ Vertical Integration

Selected Works

1963. Selling expense as a barrier to entry. *Quarterly Journal of Economics* 77: 112–128.

1964. *The economics of discretionary behavior: Managerial objectives in a theory of the firm*. Englewood Cliffs: Prentice-Hall.

1966. Peak load pricing and optimal capacity under indivisibility constraints. *American Economic Review* 56: 810–827.

1967a. Hierarchical control and optimum firm size. *Journal of Political Economy* 76: 123–138.

1967b (with Thomas Sargent). Social choice: A probabilistic approach. *Economic Journal* 27: 797–813.

1968a. Wage rates as a barrier to entry: The Pennington case in perspective. *Quarterly Journal of Economics* 82: 85–116.

1968b. Economics as an antitrust defense: The welfare tradeoffs. *American Economic Review* 58: 18–35.

1970. *Corporate control and business behavior: An inquiry into the effects of organization form on enterprise behavior*. Englewood Cliffs: Prentice-Hall.

1971. The vertical integration of production: Market failure considerations. *American Economic Review* 61: 112–123.

1975. *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

1976. Franchise bidding for natural monopolies – in general and with respect to CATV. *Bell Journal of Economics* 7: 73–104.

1977. Predatory pricing: A strategic and welfare analysis. *Yale Law Journal* 87: 284–340.

1979. Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics* 22: 233–261.

1983. Credible commitments: Using hostages to support exchange. *American Economic Review* 73: 519–540.

1985. *The economic institutions of capitalism: Firms, markets, relational contracting*. New York: Free Press.
1988a. The logic of economic organization. *Journal of Law, Economics, and Organization* 4: 65–93.
1988b. Corporate finance and corporate governance. *Journal of Finance* 43: 567–591.
1989. Transaction cost economics. In *Handbook of industrial organization*, ed. R. Schmalensee and R. Willig. New York: North-Holland.
1991. Comparative economic organization: the analysis of discrete structural alternatives. *Administrative Science Quarterly* 36: 269–296.
1995. Transaction cost economics and the evolving science of organization. In *The makers of modern economics*, vol. II, ed. A. Heertje. Aldershot: Edward Elgar.
1996. *The mechanisms of governance*. Oxford: Oxford University Press.
2010. Transaction cost economics: The natural progression. *American Economic Review* 100: 673–690.

Bibliography

Alchian, A., and H. Demsetz. 1972. Production, information costs, and economic organization. *American Economic Review* 62: 777–795.
Coase, R.H. 1937. The nature of the firm. *Economica New Series* 4: 386–405.
Coase, R.H. 1960. The problem of social cost. *Journal of Law and Economics* 3: 1–44.
Demsetz, H. 1968. Why regulate utilities? *Journal of Law and Economics* 11: 55–66.
Fischer, S. 1977. Long-term contracting, sticky prices, and monetary policy: Comment. *Journal of Monetary Economics* 3: 317–324.
Macher, J.T., and B.D. Richman. 2008. Transaction cost economics: An assessment of empirical research in the social sciences. *Business and Politics* 10: 1–63.
Masten, S.E. 1996. Oliver E. Williamson. In *American economists of the late twentieth century*, ed. W.-J. Samuels. Cheltenham: Edward Elgar.
Simon, H.A. 1957. *Models of man, social and rational*. New York: Wiley.
Simon, H.A. 1961. *Administrative behavior*, 2nd ed. New York: MacMillan.
Simon, H.A. 1991. Organizations and markets. *Journal of Economic Perspectives* 5: 25–44.

Wilson, Edwin Bidwell (1879–1964)

Murray Milgate

While primarily a mathematician, Wilson's relatively few contributions to economics in the interwar years, particularly two short papers on demand theory (1935 and 1939) and one on business cycles (1934), were not without their influence in Harvard economic circles of the day. Schumpeter drew on the arguments of Wilson's paper on the periodicity of US business cycles in his *Business Cycles,* and Samuelson's *Foundations* contains an acknowledgement to Wilson (with Schumpter and Leontief) in its preface, and credits him with the suggestion of utilizing the Le Chatelier principle in economic analysis.

The essay on cyclical fluctuations in business activity was an attempt to make deeper analytical use of the monthly index of US business activity prepared by Leonard Ayers and published in 1931. Using the device of the 'periodogram', invented by Arthur Schuster, Wilson is able to extract from Ayers' data 'hidden' cycles of different periodicities. The idea that behind any given aggregative series there might lurk different patterns of cyclical movement was, no doubt, a spur to Schumpeter's consideration of the simultaneous operation of Juglar, Kitchin and Kondratieff cycles in *Business Cycles*.

The two short essays on demand theory (1935 and 1939) are concerned with the derivation of the law of demand – that is, the inverse relationship between price and quantity demanded. The first generalizes Pareto's proof of the proposition, which had assumed additively separable utility functions. Wilson assumes instead only that $U_i(x_1 \ldots x_n)$, may take the form $U_i(x_1) + U_i(x_2 \ldots x_n)$ and derives from this the law of demand. The second paper is designed to show that Marshall's assumption of a constant marginal utility of money gave only a special case of the law of demand, and that the same result could be obtained without

it. As Wilson observed, this 'forces us over from the "index of ophelimity" to a utility definite except for a linear transformation, i.e., except for scale and origin' (1939, p. 649). The importance of this result, especially given its relation to the Hicks–Allen theory of demand, for the subsequent debates over cardinal versus ordinal utility is readily apparent.

Wilson was born at Hartford, Connecticut, on 25 April 1879. After graduating from Harvard in 1899, he took his Ph.D. from Yale in 1901. From 1907 until 1922 he was on the faculty at MIT, first as professor of mathematics and later as professor of mathematical physics. From that date until his retirement, he was professor of vital statistics at the Harvard School of Public Health. He served as president of the American Statistical Association (1929), was vice-president of the National Academy of Sciences (1949–53), and was an honorary fellow of the Royal Statistical Society. He died on 28 December 1964.

Selected Works

1912. *Advanced calculus*. Boston/New York: Ginn & Co.
1934. The periodogram and American business activity. *Quarterly Journal of Economics* 48: 375–417.
1935. Generalization of Pareto's demand theorem. *Quarterly Journal of Economics* 49: 715–717.
1939. Pareto versus Marshall. *Quarterly Journal of Economics* 53: 645–650.

Wilson, James (1805–1860)

Asa Briggs

Politician, political economist, founder and owner of *The Economist* and father-in-law of Walter Bagehot, James Wilson was born at Hawick, Scotland, the son of a millowner. After a personal financial crisis in 1837 Wilson turned to pamphleteering against the Corn Laws, which in 1839 he claimed benefited the agricultural interest no more than manufacturers or workers. In two later pamphlets, 1840 and 1842, he traced business fluctuations to the artificial influence of the corn laws and advised increased direct taxation and reduced customs and excise duties to restore prosperity. Until Sir Robert Peel, following the main thrust of this policy, repealed the Corn Laws in 1846, Wilson worked closely with Richard Cobden and the Anti-Corn Law League. *The Economist*, the first number of which, written mainly by Wilson, appeared on 2 September 1843, was a free-trade advocate which soon attracted a regular business readership as an internationally known journal of fact and opinion. In 1847, when Wilson was returned to Parliament, he published 'Capital, Currency and Banking', pleading for a 'sound currency' and opposing sections of Peel's Bank Charter Act of 1844. He also argued for the repeal of the Navigation Laws. Soon given government office in 1848, Wilson was an able Financial Secretary to the Treasury from 1853 to 1858; and in 1859, after briefly holding the Vice-Presidency of the Board of Trade, he served in India as first financial member of the Viceroy's Council with the task of reforming finances. His 1860 budget introduced a controversial income tax and later in the year, just before his death, he established a paper currency.

Wilson, Thomas (1525–1581)

Henry W. Spiegel

English lawyer and man of letters, Wilson is remembered in the history of economics for his *Discourse on Usury*, published in 1572. Wilson lived in England during his youth and early

manhood. After the accession of Queen Mary I, Wilson, a Protestant, went into exile on the Continent, where he earned a doctorate in civil law. On his return in 1560 he held a number of high offices as a Member of Parliament, judge, diplomat and Secretary of State.

Wilson wrote the *Discourse on Usury* at a time when the public attitude to interest-taking underwent a profound change. During the Middle Ages the usury rule, which outlawed interest, had prevailed. After the break with Rome, legislation was passed in England in 1546 which allowed interest up to a rate of 10%. After some wavering the legislation of 1546 was confirmed in 1571. When the latter bill was passed in Parliament, Wilson was one of the two members who voted against it; he was part of a minority then on its way out. His *Discourse* contains a belated argument in favour of the medieval prohibition of interest, an argument that he supported with copious references to the Scholastic literature. He gave a respectful hearing to the divergent views of such reformers as Calvin and Bucer, but felt unable to accept them in view of the condemnation of interest in the Bible.

Wilson's opposition to interest cannot be explained by a lack of wordly wisdom on his part. His career attests to his familiarity with the world of affairs, but his strong religious and moral convictions and aversion to the hustle and bustle of commerce made him feel not at home in this world.

The *Discourse* was reprinted in 1926, edited with an historical introduction by R.H. Tawney (London: Bell). Tawney's noteworthy introduction of some 170 pages supplies further detail about Wilson's career and his work.

Withers, Hartley (1867–1950)

Murray Milgate and Alastair Levy

British financial journalist and editor of the *Economist* from 1916 to 1921, Withers was born at Liverpool on 15 July 1867. After Westminster School and Christ Church, Oxford, he joined the staff of *The Times* in 1894 rising to become head of its City office in 1905. In 1910 he took over the City editorship of the *Morning Post*, but in 1911 moved to a position with a company of merchant bankers in the City of London. During World War I he returned to journalism at the *Economist*. These close links with the financial sector led him to colour most of his writings with a rather rosy hue when it came to assessing the role of the City of London in promoting national and international economic development. For example, in his first book, *The Meaning of Money* (1909), he concluded that 'a credit system has thus been evolved of extraordinary elasticity and perfection, so perfect in fact that its perfection is its only weakness' (pp. 295–6). In *International Finance* (1916) he professed a 'weakness for financiers' (p. 94), defending them, in particular, against charges made by the British socialist politician Philip Snowden that their interests were in war not peace and, more generally, against the traditional socialist challenge to the influence of finance capital in the economy as a whole. This argument was extended in his *Case for Capitalism* (1920), where he claimed that the system of private property and private enterprise was necessary for civilization.

In so far as his numerous books treated matters of economic theory, probably the most interesting fact is that Withers was an early proponent of what was to become known as the Cambridge cash balance approach to the quantity theory of money. He also drew the practical distinction between the activities of saving and finance, but failed to draw the analytical distinction between them which was to be so important in Keynes's work in the late 1920s and early 1930s. Although the finance of industry and trade was seen by Withers to be essentially an independent activity, he maintained that it was limited by available capital, itself accumulated by 'the quiet, prosaic, and often rather mean and timorous people who have saved their money for a rainy day' (1916, p. 83). Withers held to the traditional idea that there was a necessary

trade-off between consumption and capital accumulation (investment), and invoked this to dismiss what he called 'socialistic' claims to the effect that under a different system of economic organization and management both consumption and investment might be expanded simultaneously.

It would be incorrect, however, to view Withers as being entirely uncritical of the financial system. He was, for example, alert to what he felt to be the dangers of financial instability in the world economy; a factor which led to his advocacy of world peace and his vigorous support for the League of Nations, whose efforts to secure international cooperation rather than conflict he saw as being the only constructive international economic policy (1918). Moreover, and perhaps simply illustrating that there is nothing new under the sun, Withers was concerned with the possible consequences of any large expansion of international indebtedness. When loans went to economies whose capacity to repay was not properly assessed, there was a very real danger of default and so international monetary instability. Withers illustrated his point with the case of British loans to Honduras in the 19th century, and international loans to Brazil and Mexico in the early 20th century. Discussing phenomena which some economists of the 1970s and 1980s thought to be entirely new in their generation, Withers (1916) noted the fragility of the financial position of borrowers whose principal source of export earnings relied on primary products whose prices in international markets were subject to wide variation. Withers attributed Brazilian problems to recent dramatic falls in its then staple export, rubber. The case of the Mexican default, with its then recent revolution, arose (according to Withers) from taking insufficient account of the factor of political stability in international lending. He also discussed what is now commonly called re-scheduling of debt, insisting on terms (so familiar to debtor nations of the late 20th century) which ensured budgetary stringency at home. He opposed lending designed simply to finance domestic deficits.

His last work, *The Defeat of Poverty* (1939), was a contribution to the literature of recovery from depression. He died at Colchester on 21 March 1950.

Selected Works

1909. *The meaning of money.* London: Smith, Elder & Co.
1910. *Stocks and shares*. London: Smith, Elder & Co.
1913. *Money changing*. London: Smith, Elder & Co.
1914. *Poverty and waste*. London: Smith, Elder & Co.
1915. *War and lombard street*. London: Smith, Elder & Co.
1916. *International finance*. London: Smith, Elder & Co.
1918. *The League of nations: Its economic aspect*. Oxford: Oxford University Press.
1920. *The case for capitalism*. London: Eveleigh Nash.
1939. *The defeat of poverty.* London: Jonathan Cape.

Witte, Edwin Emil (1887–1960)

W. Cohen

Keywords
Witte, E. E.; Social security in the United States

JEL Classifications
B31

Witte was Professor of Economics at the University of Wisconsin (1933–57), President of the American Economic Association (1955), first President of the Industrial Relations Research Association (1948),

and Chief of the Wisconsin Legislative Reference Service (1922–33). His primary field was labour and social legislation. A student of John R. Commons, he was an institutional economist and a pragmatic social reformer.

His outstanding contribution was his significant role as Executive Director of President Franklin D. Roosevelt's Cabinet Committee on Economic Security (1934–5) which drafted the legislation that became the Social Security Act of 1935. Witte prepared the Committee's report and recommendations to the President. His *The Development of the Social Security Act* (1936) recounting the legislative history of the Act is an outstanding model of its type.

Witte published The *Government in Labor Disputes* (1932), assisted in the formulation of the Norris–LaGuardia Act (1932) restricting injunctions in labour disputes and he was Regional Director of the War Labour Board in Detroit (1942–5).

Witte received his Ph.D. in economics from the University of Wisconsin (1927) and was Chairman of the Department of Economics (1936–41 and 1946–57). Except for temporary assignments, he lived in Wisconsin all his life. He had a practical outlook on economic and political issues, coloured by LaFollette progressive populism.

Selected Works

1932. The g*overnment in labour disputes*. New York/London: McGraw-Hill.

1936. *The development of the social security act: A memorandum on the history of the Committee on economic security and the drafting and legislative history of the social security act*. Washington, DC. Reprinted, with a foreword by F. Perkins, Madison: University of Wisconsin Press, 1962.

Bibliography

Anon. 1960. Edwin E. Witte (1887–1960). Father of social security. *Industrial and Labor Relations Review* 14 (1): 7–9.

Schlabach, T.F. 1969. *Edwin E. Witte: Cautious reformer*. Madison: State Historical Society of Wisconsin.

Wold, Herman O.A. (1908)

E. J. Hannan

JEL Classifications
B31

Wold was born at Skien in Norway on 25 December 1908 but the family migrated to Sweden in 1912. His university education was at Stockholm, where he took his doctoral degree in 1938, under H.A. Cramér. He became Professor of Statistics at Uppsala in 1942, moved to Gothenburg in 1970, and retired in 1975.

His doctoral thesis (Wold 1934) dealt with the theory of stationary time series. Two theorems first proved in it are of lasting value. The first is the Wiener–Khintchine relation for a discrete–time series, but probably more important was the Wold Decomposition Theorem, which represents a stationary time series as the sum of an (infinite) moving average of past innovations (linear prediction errors) and a perfectly predictable component. Wold was also influential in time series analysis through his student, P. Whittle.

However, most of Wold's later work has been in econometrics. Membership of a 1938 committee to study consumer demand, rationing in case of war being the motivating force, led him to the study of general economic modelling. His work on consumer demand culminated in Wold and Juréen (1952). Economic modelling, in turn, led him to the work of Tinbergen (1939) on the statistical measurement of business cycles. Tinbergen's model was linear and connected a vector, $y(t)$, of endogenous variables to a vector, $z(t)$, of exogenous variables and lagged endogenous variables by an equation

$$y(t) = By(t) + Gz(t) + \varepsilon(t),\ E\{\varepsilon(s)\varepsilon(t)'\} = \delta_{s,t}\Omega \quad (1)$$

where the $\varepsilon(t)$ are errors. Wold observed that Tinbergen's equations were recursive in that after a rearrangement of the rows of (1) the matrix

B was lower triangular with zeros along the diagonal. If Ω is diagonal then (1) may be validly estimated by least squares and this will be the maximum likelihood method if the $\varepsilon(t)$ are also Gaussian. Wold sought to promote recursive modelling in contrast to the non-causal modelling that became influential in econometrics following Haavelmo (1944). The recursive models are causal since, after the rearrangement of rows, elements of $y(t)$ can be regarded as arranged in a causal hierarchy. Wold's view does not seem to have prevailed. The complexity of economic phenomena, including the nonlinearity of economic behaviour, the poor quality of much data, and the large amount of aggregation, together with autocorrelation of $\varepsilon(t)$, make the issue seem somewhat removed from reality.

Wold (1959) also emphasized the distinction between prediction and structural estimation and has proposed an iterative estimation of (1), oriented towards prediction, where $y(t)$ on the right is replaced by $\left(I-\widehat{B}\right)^{-1}\left(\widehat{G}z(t)\right)$, with $\widehat{B}$, $\widehat{G}$ obtained from a previous iterations.

Selected Works

1934. *A study in the analysis of stationary time series.* Stockholm: Almqvist and Wiksell.
1952. (With L. Juréen.) *Demand analysis: A study in econometrics.* New York: John Wiley.
1959. Ends and means ineconometric model building: Basic considerations reviewed. In *Probability and statistics,* The Harald Cramér volume, ed. U. Grenander. Stockholm: Almqvist and Wiksell, 355–434.
1960. *Forecasting on a scientific basis.* Lisbon: Gulbenkian Foundation.
1965. *Bibliography of time series and stochastic processes.* Edinburgh: Oliver & Boyd.
1980. *The fix-point approach to interdependent systems.* Amsterdam: North-Holland.

Bibliography

Bentzel, R., and H. Wold. 1946. On statistical demand analysis from the viewpoint of simultaneous equations. *Skand. Aktuarietidskrift* 29: 95–114.
Haavelmo, T. 1944. The probability approach in econometrics. *Econometrica* 12.
Tinbergen, J. 1939. *Statistical testing of business cycle theories, vol, II: Business cycles in the USA. 1919–9132*. Geneva: League of Nations.

Women and Work

Jane Humphries

Why is 'women and work' an issue? In the past, as in the present, most women, like most men, have worked for a living. But women's work, its particular historic development, its current content and location, and its meaning to its subjects and objects, is different from the work of men. Work is a gendered experience.

Women's work is distinguished primarily, though not exclusively, by their responsibility for certain tasks associated with daily and intergenerational reproduction. The cooking, cleaning, childcare, nursing and nurturing involved is a distinct labour in many ways: not least because it is unpaid. It has remained privatized even in advanced industrial economies whether organized by a plan or by the market, and in societies where planners' preferences or market prices index value, work which is neither planned nor marketed is undervalued, indeed often to the point of invisibility. It is not deemed to be work at all.

The differences between men's and women's work does not stop here. Paralleling the division of labour between men and women in unpaid work in the home is a division of labour by sex in paid work. Men and women are not randomly distributed across the employment structure. Rather there are men's jobs in which primarily men are engaged and women's jobs where the labour force is overwhelmingly female. There are few mixed occupations where men and women can be found in the same proportions as in the labour force as a whole.

A third significant dimension of women's jobs concerns their terms and conditions. They are less

likely to be complemented by expensive capital equipment, thus less productive, more likely to be temporary and insecure, less likely to be organized, and less associated with prospects for promotion. Above all they are less well paid.

The implications of these characteristics for the economic position of women are clearly pernicious. In the home women are unpaid and unappreciated; in the workplace their relatively low wages make them vulnerable to poverty and deny them independence from state or family subsidies which are desperately needed if there are others dependent on their wages. Their higher labour market flows and sometimes disproportionate representation in unstable jobs make them especially subject to recession-enforced periods of nonparticipation or unemployment.

Massive empirical evidence documents these characteristics of women's work as universals within the advanced industrial world (OECD 1985; ILO 1985). There are outliers. Modern Sweden, the USSR, and most countries during wartime have achieved extraordinary female involvement in paid production. Ireland is out at the opposite extreme. Eastern European countries have less sexual segmentation and different jobs are feminized, but a sexual division of labour exists nonetheless. The deployment of women in some countries has to be seen as mediated by religion, but this usually compounds rather than mitigates the privatization, segmentation and subordination suggested here as universals.

The history of women's work can also be characterized in terms of stylized facts. Thus the painful and problematic transition of women from production for use to waged work, closely documented in the British case, can and has been paralleled in other national experiences (Pinchbeck 1930; Tilly and Scott 1978). Even countries at very different levels of development can be interpreted as exhibiting aspects of the same sexual divisions of labour if the comparison is with the historical experience of the now economically advanced countries.

So overwhelming is the evidence of a *generic* women's experience of work in both the past and the present that it has prompted explanations in terms of some universal, cross–cultural, historical model of female subordination often described as patriarchy. This and other sometimes integrated universalist explanations appealing to biological or psychological differences between the sexes constitutes the first explanatory framework reviewed below.

Models of Explanation

Patriarchy

Patriarchal models of explanation see male dominance and female subordination as an enduring characteristic of all societies, hence readily explaining the commonalities of experience. An important move here is the detachment of patriarchy from other class systems of dominance and subordination and the denial, in opposition to classical marxism, of the former's dependence logically and historically on private property (Delphy 1977; Hartmann 1976).

Much of the empirical debate has been anthropological, involving the operational definition of patriarchy and its identification in preclass societies. One question of interest here is whether a sexual division of labour in work itself constitutes patriarchy, or if separate but equal roles for men and women are feasible.

The enormous changes that have taken place in women's political, legal and economic status suggest that an immutable patriarchy is indefensible, and most recent presentations within this framework argue that while patriarchal power relations remain a constant their particular expressions, and perhaps their intensity, change with economic and social development. Much of this literature is concerned to specify the form of interaction between patriarchal power relations and class society while retaining the essential autonomy of male authority and control. Adherents of this model of explanation must argue that the subordination of women in advanced capitalism is explained not by the dominant mode of production but by patriarchal power relations operating through the family and the political system and within the social relations of production. Capitalism may exploit the divisions among the working class arising from patriarchy, but above all

capitalism must adapt itself to a given system of sexual hierarchy. Thus the conditions of women's employment are primarily determined by the dominance of men and capital must adapt to a sex-differentiated wage hierarchy so that men's power in the domestic and political spheres is not contradicted in the workplace. Adaptation takes place despite the interests of capital in eroding male power and establishing a homogeneous competitive labour force. Alleged mechanisms of control have received attention: ideology and socialization as well as concrete institutions like schools and trade unions.

One of the major criticisms of patriarchy as an explanatory framework is that it contains no material explanation of women's position and ultimately the argument must devolve on biologism, psychoanalytic theory or cultural catalepsy. Another possibility is to see patriarchy itself as an economic class system with men as the appropriators of female labour-time: a prospectively rich but empirically underdeveloped approach that is unfortunately thwarted by the suspicion that much of women's potentially alienable labour time benefits children rather than fathers.

Whatever the origins of patriarchy, they are clearly independent of the capitalist mode of production. There is therefore truth in the argument that women's economic subordination has its roots in precapitalist forms and cannot be explained solely by functionalist reference to capitalism. The third conceptual framework reviewed below owes much to this challenge to classical marxism.

Neoclassical Economics

Neoclassical economic theory tries to explain women's economic subordination in terms of rational utility-maximizing behaviour. The emphasis has been on the narrowly economic issues of women's labour supply, occupational distribution and relative pay. Neoclassical analysis of labour supply postulates an individual allocating his/her time between work and leisure. Leisure here is a catch-all term for all uses of time other than paid work but it is a peculiarly male label in view of the reality of domestic labour and the latter's impact on women's 'leisure'. Indeed, much of the neoclassical interest in female labour supply has occurred precisely because the question 'Why do paid work?' can meaningfully be asked of women, for they have the option of work in the home. The added dimension to women's choice set challenged neoclassical economists, who soon realized that female labour supply could not be modelled analogously to that of men and significant innovations with widespread repercussions for mainstream economics followed. Two relevant developments are sketched below.

First, recognition that simple reduced form models of female labour supply, drawing only on samples of working women, involved serious biases, prompted the development of structural models in which labour supply decisions were modelled as involving discrete quantitative choices about whether or not to work and how many hours to work. These second generation labour-supply models involved new techniques for handling discrete choices (logit, probit and tobit) and the introduction of the shadow or imputed wage to capture the influence of unobservable but relevant returns, for example, to non-market activities (Heckman 1974; Heckman et al. 1981; Killingsworth 1983).

As longitudinal data have become increasingly available, particularly in the United States, investigation has focused on lifetime experience and appropriate techniques for handling analyses of attachment to the labour force over the whole life cycle have been developed. Much empirical work has now been undertaken for the USA and UK providing important information regarding female labour supply.

However, the widespread use of personal characteristics and family circumstances as proxies for the tastes of individuals or the shadow wage of domestic work, indicates the limitations of the approach (Greenhalgh and Mayhew 1981). Belief that these variables genuinely represent exogenous tastes or comparative advantages surely involves some dubious propositions about sex and race-linked biological endowments! However, the attachment of certain behavioural proclivities to variables like race, sex and the

presence of small children may simply reflect rational expectations about discrimination or family opposition to waged work, or even historical differences in the relative earnings potential of men and women. Then labour supply cannot be taken as independent of demand-side variables or the organization of the family and a historic and interactive analysis between production and social reproduction must be undertaken.

A second development stemmed from recognition that husbands' and wives' decisions about work, leisure and homeproduction are interdependent, which contested the basis of neoclassical economic theory in the individual decision-maker. Simultaneously neoclassical economists, looking for new areas of behaviour to subject to their choice theoretic framework, lighted on fruitful terrain: the household. Becker's (1965) analysis of household decisionmaking in terms of his theory of the allocation of time produced the New Household Economics (NHE).

The NHE views the household as a production unit and consumption decisions as dictated by the drive to maximize utility. Various activities can contribute to utility and these activities require inputs of time and other goods. The ultimate product is the utility derived. All kinds of household decisions from the mundane to the consideration of whether to have children of certain qualities and quantities, have been cast in this framework. Economists' techniques can then be mobilized to describe the optimal allocation of time under certain assumptions about tastes and relative costs.

The value of the NHE lies in its explication of the link between labour supply decisions and consumption decisions and therefore between women's primary responsibility for work in the home and their partial and discontinuous involvement in paid work. Unfortunately the insight is not maintained. Other relevant problems are sidestepped or trivialized: for example, the deep difficulties involved in synthesizing a collective preference ordering from individuals' preferences are not addressed here despite the move from the individual to the household. One way out is to postulate a set of rules for aggregating and weighting individuals' preferences; another is to understand the collectivity's preference ordering as 'given' by an *individual* representative: 'the benevolent dictator' of welfare economics. Both methods abstract from the conflicts and complementarities among household members and, essentially, collapse household and individual decision-making. Moreover, feminists see the benevolent dictator as mirroring the dominant patriarchal form of family organization in society and, not surprisingly, are leery of closing the model by accepting the very hierarchy of authority that they want to question. Perhaps the recent interest in bargaining models of family relations (Pollak 1985) will eventually help with these difficulties.

More importantly, the NHE cannot explain why the most efficient allocation of time by family members should involve a sexual division of labour between paid and unpaid work. Specialization is explained by comparative advantage, but the latter's suggested origins often make the argument circular: women hiring men as breadwinners because they earn more, but women earning less because they leave the labour market to have children (Becker 1981). Alternatively the argument is sometimes shored up here by biologism: higher productivity in domestic labour simply being read off the female sex, or women's prior investment in children, since they carry them in the womb, being held to make them more inclined to further childcare investment (ibid). So comparative advantage is deduced from the existing sexual division of labour and then used to explain that division.

Neoclassical economics' treatment of the occupational distribution of women workers and their relative terms and conditions, suffers from similar defects. Briefly both occupational choice and relative wages are viewed as the outcomes of rational utility–maximizing behaviour. Indeed the former is often held as an 'explanation' of the latter as women's disproportionate representation in poorer paying jobs accounts for a substantial part of the male–female wage differential. Sometimes occupational choice and relative wages are analysed in the framework which subsumes them both into sex specific choices about levels of investment in human capital. Employers' discrimination has usually entered the argument only

to explain the residual after the impact of other sex specific worker characteristics have been deducted.

Why do women choose certain occupations in preference to alternatives despite the fact that the resulting occupational distribution is (demonstrably) a major factor in their relatively low pay and poor terms of employment? An obvious defensive manoeuvre, and one which neoclassical economists have not ignored, despite its disturbing implications for the premise of rational-maximizing behaviour, is to cite pre–market discrimination as engineering women's choices, or cultural predispositions to study certain subjects as leading girls into less well-paid jobs. But these arguments are *non sequiturs* as far as economic explanations of women's subordination are concerned.

Human capital theory has also been used to explain such seemingly paradoxical choices as rational-maximizing behaviour. Polachek (1975, 1976) has argued that women *choose* occupations for which earnings losses during spells of non-participation are minimized. Since women plan intermittent paid work because of intended childrearing, they prefer occupations where skills depreciate only slowly when not employed. Women's primary responsibility for children is assumed in this model, as are differences in the rates of decay of human capital across occupations. Polachek's work has been subject to both theoretical and empirical criticism internal to his own paradigm (England 1982; Beller 1982). Radical reservations have also been expressed about the notion of productivity embedded in human capital theory. It is essentially a sexist concept since it only counts as productive those skills which the market rewards and many skills which women have go unrecognized (Dex 1985).

To summarize: neoclassical economic theory confuses *descriptions* with *explanations* of the subordination of women. Moreover, cultural and even biological factors have sometimes been used as prime movers in the argument, despite the inconsistency of this procedure with a belief in the prime explanatory power of rational maximizing behaviour. This is not to deprecate the value of much neoclassical work on female labour supply, or the insights of the NHE, but as a model of explanation of either women's specialization in the home or the sexual division of labour in paid work, the approach is ultimately nugatory.

Neo-Marxist Economics

The neo-marxist approach developed out of the challenge to classical marxism's treatment of women's subordination posed by the persistence of patriarchal social relations. Interest was initially focussed on the articulation of unpaid work in the home to production relations in a capitalist economy and the meaning of the former for value categories: the domestic labour debate (Dalla Costa and James 1972). How domestic labour should be integrated into value accounting represented a challenge to marxian value theory and to classical analyses of the natural price of labour (Himmelweit and Mohun 1977). The domestic labour debate's attention to work done to reproduce workers and their labour power promoted new interest in the process of *social reproduction* more broadly defined (Kuhn and Wolpe 1978). Consequently developments in the theory of the family have both drawn on and contributed to developments in the theory of the state Both have contributed to significant progress in neo-marxist understanding of advanced capitalism. It is interesting to note that neo-marxist interest too is now turning to bargaining models of the family.

Attention also spread to analyses of the relationship between women's primary responsibility for the reproductive work of the home and their position in paid labour. Concepts for classical marxism, such as the reserve army of labour, as well as from contemporary labour economics, such as labour market segmentation, have been especially useful in analyses of the impact of secular restructuring and the business cycle on women workers. Recent studies have tried to test whether women do constitute a buffer labour reserve over the business cycle or if their segregation in the less volatile sectors and occupations affords them relative protection in hard times (Bruegel 1979). A third hypothesis which has received attention is whether women's cheapness in conjunction with their predominance in growing sectors is causing women to be substituted cyclically and secularly for men (Rubery 1987).

These studies have developed a dual systems theoretic approach which sees women's economic subordination as the outcome of interaction between social reproduction and production. Although debate continues as to the precise specification of this interaction, key characteristics of the neo-marxist approach to understanding women's work include: (1) an insistence that social reproduction be taken seriously as work and as an integral part of the economy; (2) a conviction that social reproduction is relatively independent of the organization of production and does not respond smoothly, accommodatingly or predictably to the needs of the economy; (3) the deduction therefore that the relationship between the spheres of production and social reproduction can only be understood historically and is not predetermined; and (4) also that the relationship must be analysed within a nonfunctionalist perspective (Humphries and Rubery 1984).

The approach is less methodologically hidebound and much more open-ended than that of neoclassical theory. Nevertheless developments within this literature have both drawn from and fed back into neoclassical economics. Moreover both approaches are forced to respond to the contemporary dramatic increases in female participation rates in advanced industrial economies and attempt to predict the implications for the unequal burden of domestic work, for occupational segregation, and for the terms and conditions of women's work.

Conclusion

As suggested above, interest in women's work is not only analytical. Social commentators of the past, as well as of the present, have attributed major significance to shifts in the allocation of women's labour time between the home and the workplace, though not always in agreement as to their implications. Contrast the oft-quoted view of Engels that 'The emancipation of women will only be possible when women can take part in production on a large social scale, and domestic work no longer claims anything but an insignificant amount of her time' (1891, p. 221) with the more traditional position of Marshall: 'If we compare one country of the civilized world with another, or one part of England with another, or one trade . . . with another, we find that the degradation of the working classes varies almost uniformly with the amount of rough work done by women' (1961, p. 565).

Perhaps the reconciliation of these two distinguished views lies in consideration of the terms and conditions of women's paid work and how it is coordinated with childcare and domestic labour. Significantly it is college-educated women with their more interesting, betterpaid jobs, who are able to purchase domestic help and high-quality childcare, who find work most enriching. For women as for men, one objective must be more interesting work. More contentious, in terms of the distribution of necessary labour time between the sexes, is the target of a more equal distribution of domestic responsibilities.

See Also

- ▶ Gender
- ▶ Labour Market Discrimination

Bibliography

Becker, G. 1965. A theory of the allocation of time. *Economic Journal* 75(September): 493–517.

Becker, G. 1981. *A treatise on the family*. Cambridge, MA: Harvard University Press.

Beller, A. 1982. Occupational segregation by sex: Determinants and changes. *Journal of Human Resources* 17(3, Fall): 371–392.

Bruegel, I. 1979. Women as a reserve army of labour: A note on recent British experience. *Feminist Review* 3(Autumn): 12–23.

Dalla Costa, M., and S. James. 1972. *The power of women and the subversion of the community*. Bristol: Falling Wall Press.

Delphy, C. 1977. *The main enemy: A materialist analysis of women's oppression*. London: Women's Research and Resources Centre.

Dex, S. 1985. *The sexual division of work*. Brighton: Wheatsheaf.

Engels, F. 1891. *The origin of the family, private property and the state*, 4th ed. New York: International Publishers. 1972.

England, P. 1982. The failure of human capital theory to explain occupational sex segregation. *Journal of Human Resources* 17(3, Fall): 358–370.

Greenhalgh, C., and K. Mayhew. 1981. Labour supply in Great Britain: Theory and evidence. In *The economics of the labour market*, ed. Z. Hornstein, J. Grice, and A. Webb. London: HMSO.
Hartmann, H. 1976. Capitalism, patriarchy and job segregation by sex. *Signs: Journal of Women in Culture and Society* 1(3), Pt 2, Spring: 137–169.
Heckman, J. 1974. Shadow prices, market wages and labor supply. *Econometrica* 42(4, July): 679–694.
Heckman, J., M. Killingsworth, and T. MacCurdy. 1981. Empirical evidence on static labour supply models. In *The economics of the labour market*, ed. Z. Hornstein, J. Grice, and A. Webb. London: HMSO.
Himmelweit, S. and S. Mohun. 1977. Domestic labour and capital. *Cambridge Journal of Economics* 1(1, March): 15–31.
Humphries, J. and J. Rubery. 1984. The reconstitution of the supply side of the labour market: the relative autonomy of social reproduction. *Cambridge Journal of Economics* 8(4, December): 331–346.
ILO. 1985. *World labour report*, vol. II. Geneva: ILO Publications.
Killingsworth, M. 1983. *Labour supply*. Cambridge: Cambridge University Press.
Kuhn, A., and A. Wolpe (eds.). 1978. *Feminism and materialism*. London: Routledge & Kegan Paul.
Marshall, A. 1961. *Principles of economics*, 9th ed. London: Macmillan.
OECD. 1985. *The integration of women into the economy*. Paris: OECD.
Pinchbeck, I. 1930. *Women workers in the industrial revolution, 1750–1850*. London: Frank Cass. 1969.
Polachek, S. 1975. Discontinuities in labour force participation and its effects on women's earnings. In *Sex discrimination and the division of labor*, ed. C. Lloyd. New York: Columbia University Press.
Polachek, S. 1976. Occupational segregation: An alternative hypothesis. *Journal of Contemporary Business Studies* 5(1): 1–12.
Pollak, R. 1985. A transaction cost approach to families and households. *Journal of Economic Literature* 23(2, June): 581–608.
Rubery, J. (ed.). 1987. *Women and recession*. London: Routledge & Kegan Paul.
Tilly, L., and J. Scott. 1978. *Women, work and family*. New York: Holt, Rinehart & Winston.

Women's Wages

J. Rubery

Women's average wages are consistently lower than men's average wages in all countries, even after adjustments for differences in working hours. These lower wages cannot be simply explained by differences in the productivity of women workers, or by the segregation of women into different jobs: they are related to the role of women in the social reproduction sphere, that is to their expected contributions to domestic labour and to family income. However, women's wages should not be identified as a separate issue; to do so suggests that it is women's wages that do not conform to a competitive norm and therefore require separate analysis as an anomaly. Women form too large a segment of the labour force for this 'anomaly' not to affect the other segment, 'male labour', and men's role in social reproduction has an equal and specific impact on their characteristics as wage labour. There is nevertheless an argument on social and political grounds for singling out women's wages for special study. Women's wages are not only low at the average or macro level, but also are consistently lower than men's at the micro level of the occupation, firm or industry. Women account for overwhelmingly the largest share of low-paid adult workers in the UK, so that ten years after the Equal Pay Act it is still reasonable to talk of a separate set of wages for women to that available to the majority of men.

Neoclassical Explanation of Women's Wages

There are three different types of hypotheses that have been put forward within a neoclassical framework to explain women's lower wages. The first is the wage discrimination hypothesis, associated with Becker's (1971) work on racial discrimination, under which women are paid less than their marginal products to compensate either employers or co-workers for their distaste for female workers. The second and third hypotheses assume that women are paid relative to their actual marginal products; according to the second hypothesis women are less productive employees than male employees, and under the third hypothesis, women are employed in less productive jobs than men, but not necessarily because they are inherently less productive workers. The

discrimination hypothesis of Becker was found to be rather difficult to reconcile with neoclassical theory unless one could assume all employers were equally discriminatory. A less or non-discriminating employer would be able to compete successfully against the established group of discriminating employers.

The second hypothesis is associated with human capital theory; women's lower wages are attributed to lower levels of educational training and perhaps more significantly, to their lack of continuous work experience which develops skills and also renews and updates them. Pre-market discrimination, in-market discrimination or personal and household preferences could all account for these different patterns of human capital acquisition, but whatever the cause the lower wages are taken to result directly from women's lack of skills relative to men's. Women's role as wives and mothers may have an indirect effect on productivity through lowering human capital acquisition, but there may also be a direct effect if women take prime responsibility for the family and they behave in ways which make them less committed and less productive workers, such as working only part-time hours or having a higher tendency to absenteeism or to voluntary quits.

One of the consequences of such characteristics might be that women are in fact confined to lower-productivity jobs, because of the difficulties of adjusting all jobs to meet different behaviour patterns and characteristics of workers. This 'confinement' may be a demand side phenomenon, employers making assumptions about women's behaviour patterns, or a supplyside phenomenon, women choosing jobs which allow them to carry out their domestic responsibilities. The origins of these different preferences or 'tastes' are taken either to be exogenous to the economy and outside the sphere of economics, or as in the New Household Economics, to arise out of the process of welfare maximization for the household, instead of for the individual. Thus one explanation of the third hypothesis, job segregation, is women's role in social reproduction.

Segregation may also be held to result because of either prejudice, based on custom and practice, about which jobs are suitable for women, or thirdly, because women have essentially different attributes or skills to those of men. All three versions of the job segregation hypothesis are compatible with a view that the jobs women do are low skilled, low productivity jobs, but with the latter two versions it could be the over-supply of women to a relatively narrow range of jobs which results in lower supply prices, more labour-intensive technology and consequently lower-marginal products (Mill 1848; Bergmann 1971). According to these two versions if demand for labour in these types of jobs rose, women's wages would be expected to rise accordingly, but under the first social reproduction hypothesis, an increase in demand for labour in these type of jobs could stimulate substitution of women by more 'committed' and 'productive' workers. The hypothesis of job segregation through prejudice is open to the same objections as were raised against Becker's hypothesis; unless the basis for segregation is real differences in skills then it would be broken down eventually by non-prejudiced employers. The 'economics of information costs' has helped to restore this hypothesis by suggesting that it may be rational for employers to use cheap 'screens' such as sex in their recruitment decisions, to avoid hiring and firing costs. Women may be excluded from a segment of the labour market either because on average women have less desirable characteristics than men (Phelps 1972), or because no women think it worthwhile to acquire the skills for entry into this segment, so the employers' beliefs remain untested (Spence 1973).

It is, nevertheless, more comfortable for neoclassical economists to attribute low wages for women to supply-side characteristics, to differences in attributes or preferences arising out of biological or social and cultural factors, than to place the burden of explanation on demand-side imperfections which prevent the equalization of returns to productivity. Under neoclassical analysis the forces of competition will always be working towards undermining these demandside constraints, but as economists offer no analysis of the forces of social and biological change, persistence of inequality can be more readily accounted for, and changes in the economic status

of women can be attributed to exogenous changes in tastes. However it is this procedure that also reveals the weakness of neoclassical analysis (Humphries and Rubery 1984); by eschewing the need to develop an analysis of the historical relationship between the organization of production and the organization of social reproduction, neoclassical theorists make adjustments to the preference functions in their models, not to take account of changes they have identified in the organization of social reproduction but to find a better fit for their model when the previous preference function fails to perform adequately. There is thus no theoretical basis for making these changes (Tarling 1981). This critique of neoclassical methodology suggests that it may be more appropriate to consider women's preferences in the labour market as conditioned by past historical experience and responses to current opportunities than as an independent cause of low wages.

A second major problem with the neoclassical analysis of women's wages arises from its assumptions that relative wages reflect relative marginal products, so that the issue to be explained is why women workers have low marginal products. Clearly, therefore, all the critiques which apply to the marginal productivity approach to wage determination also apply here to the specific issues of women's wages. However this issue also highlights some of the deficiencies of the neoclassical approach. For example, the analytical framework is based on competitive wage determination, with market clearing, but historical and cross cultural empirical evidence suggests that the female wage labour market is rarely if ever cleared. Surveys indicate a high level of hidden unemployment amongst economically inactive women, so that there are still large supplies of labour available at current wage rates. Secondly, the neoclassical approach implies that wage relativities should reflect relative skills and relative efficiencies of labour. Women's wages are in practice remarkably uniform, displaying much lower dispersion at micro and macro levels than male wages. The exclusion of women from more skilled work only provides a partial explanation, as the usual practice is to place all women's jobs whatever their characteristics, and all women, regardless of their skill or experience, within a narrow band of pay at the bottom of the pay hierarchy (Craig et al. 1985). Thirdly, if marginal products influence wages, changes in the ratio of female to male pay should come about as a result either of a change in the distribution of women within the labour market, or as a result of changes in demand or supply in the female labour market. In practice changes in the ratio have been associated, at least in Britain, more with social and institutional forces than with changes in labour market opportunities.

These considerations suggest that women's lower wages may be determined to some extent prior to the allocation of women to jobs, and independently of their characteristics and attributes. This proposition is taken up by two of the three non-neoclassical theories examined below, that is the patriarchy theory and the family wage theory. This type of approach was rejected by neoclassical theorist because it is not easily compatible with theories of competitive equilibrium. Segmentation theory, the other non-neoclassical perspective examined below, argues that there is in fact no necessary tendency for the system of competition to bring about equalization of returns to productivity for the labour employed.

Non-Neoclassical Explanations of Women's Wages

Patriarchal theories of women's wages start from the assumption that the fundamental explanation of women's inferior position in the economic system is their inferior position in the social and cultural system. Patriarchal social relations existed prior to capitalism and capitalism has had to adapt to a patriarchal system (Hartman 1979). Women earn lower wages than men in order that there should be no challenge to the system of authority within the family and the social and political structure. Employers share the patriarchal values of society, so that the system is not subject to challenge by profit-seeking entrepreneurs. Inequality in pay may be reinforced by the subordination of women in inferior or 'feminine' jobs, but the low wages do not arise out of the characteristics of the jobs

but out of the characteristics of women's position in a patriarchal society.

In contrast, the labour market segmentation approach, in its simple form, locates the causes of female inequality in the process of uneven development of the capitalist economy. Women are concentrated in particular sections of the labour market, but unlike the neoclassical model of dualism, the origins of the division between the so-called 'primary' and 'secondary' sectors are not social and institutional imperfections which distort the market, but the requirements of the economic system itself (Doeringer and Piore 1971). Primary employment sectors develop in order to maximize advantages from operating at efficient levels of capacity with a stable and fully-trained labour force. Fluctuations in demand are dealt with by subcontracting to secondary employment sectors. In addition, whole industries are located in the secondary sector if the demand for the product is generally variable and unstable. The explanation of the divergence between returns to labour in the two sectors is rooted in the operation of the economic system, but in this dual labour market model, the explanation of why women are concentrated in the secondary sector has to be looked for elsewhere. The radical version of labour market segmentation theory (Edwards et al. 1975) provides an explanation both of the structuring of the labour market into primary and secondary sectors and of the allocation of women to the secondary portion. To forestall the development of class consciousness, capitalists, it is argued, segmented or divided the labour force by creating 'artificial' hierarchies or divisions. In order to minimize the likelihood of alliances being formed across these divides, workers with different social characteristics were allocated to different segments: hence the concentration of women and ethnic minorities in secondary segments.

In the family-wage approach the analysis of women's position in the family and social system is linked directly to the analysis of the forces of production (Humphries 1977; Beechey 1978). Reliance simply on Marx cannot provide an adequate theory of women's wages because of the absence of a theory of the family. In contrast to the predictions of Engels, the nuclear family failed to 'wither away' to produce a wage labour market of undifferentiated individuals. Instead individuals on the labour market are still reproduced within a sex-differentiated social and family system and within a social reproduction system which provides forms of income support to non-wage labour. It is the differences in men's and women's relationships to the social reproduction system that leads to their labour being supplied on different terms. Men's wages are based on the cost of their own social reproduction and that of their dependents, but women's wages are based on only part of the cost of their own reproduction, on the assumption that they have access to support from either their husbands or their fathers, an assumption which is reinforced by social security systems which deny married women access to income support in their own right.

The tendency for women's labour to be supplied at below the value of labour power (that is below the average cost of reproduction) has specific consequences for the mobilization and utilization of female labour within the productive system. Female labour will tend to be mobilized at times when there are strong competitive pressures on capital to restore the falling rate of profit and female labour will tend to be concentrated in those sectors where capital is under particular pressure to force the cost of labour down below the value of labour power. Individual households do not in fact exercise 'choice' over their sexual division of labour as women are confined to jobs which offer wages below the cost of average adult subsistence. Thereby the structure of wages in the productive sphere serves to reinforce the system of social organization on which it is founded.

Under this family wage hypothesis, therefore, the lower wages that women are paid derive primarily from their own social characteristics, in particular from their position in the family income system, and not from the characteristics of the jobs that they perform. Women's low wages are thus assumed in some sense to be independent of the jobs they perform, as in the patriarchy argument. However, contrary to the simple patriarchy notion, women's lower wages are identified as having a materialist base which relates both to

their role in social reproduction and to their role in production. The lower wages are used not simply to reproduce patriarchal values but to serve competitive objectives. Thus even though low wages for women are not caused by different patterns of job allocation they may themselves lead to different patterns of female labour throughout the economy.

In order to develop a fully adequate theory of women's wages it is necessary to combine the insights of the family-wage hypothesis with the perspectives on job segregation offered respectively by the patriarchy and labour market segmentation approach. There is now a considerable body of empirical research which suggests that, at least in the UK, women's low wages cannot be explained by the characteristics of the jobs they perform. Many women's jobs require skills and experience, but these factors are not reflected in pay or grading of the jobs (West 1982; Craig et al. 1985). Moreover, work carried out within the context of the labour process debate suggests that there is no direct relationship between the skill of a job, however measured, and its pay and status. The characteristics and the bargaining power of the workers employed are more important explanatory variables. Indeed in order to 'deskill' a job it may be necessary to employ workers of lower status and bargaining power. However the substitution of women for men in order to deskill jobs has not been a universal process, so we need to understand the limits to this process. It is here that patriarchal relations may play an important role in setting up boundaries between men's work and women's work which are only breached when economic or other social forces are strong enough to break down the customary division of labour by sex. Moreover when the old division of labour is broken down and a new division established, newly feminized jobs are quickly redefined as only suitable for women, and the new division of labour is rigidified by these social values and by the continuing differentials in male and female pay.

Segmentation theory helps to explain why the utilization of women within the production sphere is concentrated in specific areas. Within this approach, it is argued that capitalism is subject to a process of uneven development, with different systems of competition prevailing in and between different sectors of the economy. As competition cannot be reduced to simply cost minimization, there is no necessary tendency for the incentives to substitution to be such as to ensure equalization of wage costs between firms or categories of labour. However, once a firm or sector becomes organized around low wage labour that cost structure becomes built into its system of organization.

The incentive to substitute low wage labour for higher paid labour is constrained by the firm's other objectives; such substitution might endanger the overall efficiency of the firm, by increasing the likelihood of instability among the experienced labour force or by reducing overall cohesion and cooperation. Moreover labour markets are also regulated by trade union organization and government employment policy which reduce firms' discretion in both wages and employment decisions. The actual form of trade union organization and government labour regulations differ considerably between countries and have a major impact both on the established norm for female wages and on the specific ways in which female labour is utilized in the economy. Thus trade union wage policies, the type of legal minimum wage system and the employment protection and benefits associated with part-time work have an impact on the terms and conditions under which women are employed which is relatively independent of the characteristics of the jobs that they perform in any particular country.

Women's wages cannot be explained solely by reference to women's family position, as the wage levels are also influenced by the system of wage determination and employment protection that prevails in the labour market. It is significant that it is countries with more egalitarian trade union wage policies or more effective minimum wage policy that tend to have a higher earnings ratio for women to men although these always stop short of unity.

These higher ratios automatically raise women's contribution to family income, and as these higher pay levels become permanently established, so the dependence of the family on

women's income is strengthened. However it is not clear whether differences in systems of family organization between countries are themselves a cause of differences in labour market earnings for women or an outcome of these differences. This issue raises important questions over the ways in which women's pay inequality could be reduced. It could be argued that as women's inferior economic status can be attributed primarily to their historical role as dependents in the family system, then it is to changes in family organization that we must look before any real progress in women's position can be made. If women were to become reliant on their own earnings for their subsistence, the 'natural price' of female labour would change in theory; however it is possible that this change would reduce the standard of living of many women unless there are mechanisms by which the necessary real wages to maintain current consumption standards could be secured in the labour market. Indeed it could be argued that increasing numbers of American women have been thrown into poverty because a change in family organization towards more single parent families has preceded the development of any effective mechanisms for women to improve upon the wage levels which relate to an outdated system of family organization. If instead gains for women are secured through trade union organization, government labour market regulations or other means, then these are likely to be relatively easily and quickly translated into a new family budget structure which then becomes a material basis for changes in women's social and family roles.

See Also

- ▶ Gender
- ▶ Inequality Between the Sexes
- ▶ Labour Supply of Women

Bibliography

Becker, G. 1971. *The economics of discrimination*, 2nd ed. Chicago: Chicago University Press.

Beechey, V. 1978. Women and production: A critical analysis of some sociological theories of women's work. In *Feminism and materialism*, ed. A. Kuhn and A.-M. Wolpe, 155–197. London: Routledge and Kegan Paul.

Bergmann, B.R. 1971. The effect on white incomes of discrimination in employment. *Journal of Political Economy* 71(2): 294–313.

Craig, C., E. Garnsey, and J. Rubery. 1985. *Payment structures in smaller firms: Women's employment in segmented labour markets*, Department of Employment Research Paper no. 48. London: Department of Employment.

Doeringer, P., and M. Piore. 1971. *Internal labour markets and manpower analysis*. Lexington: D.C. Heath.

Edwards, R., M. Reich, and D.M. Gordon (eds.). 1975. *Labour market segmentation*. Lexington: D.C. Heath.

Hartman, H. 1979. The unhappy marriage of Marxism and feminism: Towards a more progressive union. *Capital and Class* 8: 1–33.

Humphries, J. 1977. Class struggle and the persistence of the working-class family. *Cambridge Journal of Economics* 1(3): 241–258.

Humphries, J., and J. Rubery. 1984. The reconstitution of the supply-side of the labour market: The relative autonomy of social reproduction. *Cambridge Journal of Economics* 8(4): 331–346.

Mill, J.S. 1848. *Principles of political economy*, 1st ed. London.

Phelps, E.S. 1972. The statistical theory of racism and sexism. *American Economic Review* 62(4): 659–661.

Spence, A.M. 1973. Job market signalling. *Quarterly Journal of Economics* 87(3): 355–374.

Tarling, R. 1981. The relationship between employment and output: Where does segmentation theory lead us? In *The dynamics of labour market segmentation*, ed. F. Wilkinson, 281–290. London: Academic Press.

West, J. (ed.). 1982. *Work, women and the labour market*. London: Routledge & Kegan Paul.

Women's Work and Wages

Francine D. Blau and Lawrence M. Kahn

Abstract

Since the early 1980s the gender wage gap has fallen in most economically advanced countries, although a gender wage differential remains in all countries. We first document for several industrialized countries recent trends in the gender gap in labour force participation and earnings. We then outline several explanations for the gender wage gap at a given

point in time, changes in the gender gap over time, and differences in its extent across countries. Next, we consider the empirical evidence in support of various explanations. We conclude with some thoughts about future prospects for the gender wage gap.

Keywords

Affirmative action; Anti-discrimination law; Black-white labour market inequality in the United States; Centralized wage-setting; Collective bargaining; Comparable worth; Compensating differentials; Education; Gender wage gap; Human capital; Labour force selectivity; Labour market discrimination; Labour supply; Monopsony; Occupational segregation; Search costs; Skill-biased technical change; Statistical discrimination; Technical change; Trade unions; Wage differentials; Wage distribution; Wage structure; Women's work and wages

JEL Classifications

J16

Since the early 1980s the gender wage gap has fallen in most economically advanced countries, although a gender wage differential remains in all countries. While labour market outcomes for men and women may vary across a number of dimensions, economists have particularly focused on analysing gender differences in wages. This emphasis reflects a number of factors. The wage is a major determinant of economic welfare for employed individuals, as well as of the potential gain from market work for those not currently employed. Further, it affects decisions ranging from labour supply to marriage and fertility, as well as bargaining power and relative status within the family.

This article begins with an overview across a number of economically advanced countries of labour force participation and earnings differences between men and women in the labour market, and delineates the recent trends. We then consider the explanations that have been offered for the gender wage gap at a given point in time as well as for changes in the gender gap over time and differences in its extent across countries. Next, we consider the empirical evidence in support of various explanations. We conclude with some thoughts about future prospects for the gender wage gap.

Overview of Gender Differences in Labour Force Participation and Wages

Although the focus of this article is the gender wage gap, it is useful to consider the evolution of labour force participation rates by gender. Women's rising labour force participation implies that women's wage gains, discussed below, apply to an increasing share of the female population. In addition, changing female participation can mean that the qualifications and experience of the typical employed woman may be changing as well. To the extent that women's rising participation is associated with rising labour force attachment of women over the life cycle, the average level of labour market experience of women will eventually increase. Changing participation rates may also be associated with changes in labour force selectivity of women, depending on the relative qualifications of entrants and incumbents. And these factors in turn may help us explain the evolution of the gender wage gap. Table 1 shows that, across ten economically advanced countries, women's labour force participation rates rose steadily between 1979 and 2000, both absolutely and relative to men, with a faster increase in the 1980s than the 1990s. For example, taking an unweighted average of the countries listed in the table, we see that women were 64 per cent as likely as men to be in the labour force in 1979; by 1990, this ratio had risen to 77 per cent and, by 2000, it was 83 per cent. Throughout this period, Scandinavian women had especially high participation rates.

Table 2 shows female to male pay ratios from the OECD across the same ten countries shown in Table 1. The entries in Table 2 are intended to show the price of labour; in some cases the data are available as weekly or monthly earnings of full-time workers (Panel A), and in others as

Women's Work and Wages, Table 1 Labour force participation rates by gender (ages 15–64), ten Western countries, 1979–2000

	1979			1990			2000		
	Men	Women	Ratio: (women/men)	Men	Women	Ratio: (women/men)	Men	Women	Ratio: (women/men)
Australia	87.6	50.3	0.574	84.4	61.5	0.729	81.9	65.4	0.799
Finland	82.2	68.9	0.838	79.6	73.5	0.923	76.4	72.1	0.944
France	82.6	54.2	0.656	75.0	57.2	0.763	74.4	61.7	0.829
Germany	84.5	49.6	0.587	79.0	55.5	0.703	78.9	63.3	0.802
Japan	89.2	54.7	0.613	83.0	57.1	0.688	85.2	59.6	0.700
Netherlands	79.0	33.4	0.423	79.7	52.4	0.657	83.9	65.7	0.783
New Zealand	87.3	45.0	0.515	83.0	63.2	0.761	83.2	67.5	0.811
Sweden	87.9	72.8	0.828	86.7	82.5	0.952	81.2	76.4	0.941
United Kingdom	90.5	58.0	0.641	88.3	67.3	0.762	84.3	68.9	0.817
United States	85.7	58.9	0.687	85.6	67.8	0.792	83.9	70.7	0.843
Average	85.7	54.6	0.636	82.4	63.8	0.773	81.3	67.1	0.827

Notes: For 1990 and 2000, ages are 16–64 for Sweden, the United Kingdom and the United States. Germany is defined as West Germany in 1979 and 1990 and unified Germany in 2000
Sources: OECD (1990, p. 200; 2004, pp. 294 and 296)

Women's Work and Wages, Table 2 Female to male ratios, median full time earnings: ten Western countries, 1980–2000

	1980	1990	2000	Changes: 1980–2000	
				Absolute	%
A. Weekly or monthly earnings, full-time workers					
Australia	0.813	0.818	0.828	0.016	1.9
W. Germany	0.705	0.738	0.793	0.088	12.5
Japan	0.583	0.594	0.654	0.071	12.2
New Zealand	0.733	0.773	0.815	0.082	11.2
United Kingdom	0.647	0.688	0.761	0.113	17.5
United States	0.634	0.715	0.748	0.114	17.9
Average	0.686	0.721	0.766	0.081	11.8
B. Annual earnings, full-time, year-round workers					
Finland	0.734	0.771	0.796	0.062	8.4
France	0.803	0.847	0.905	0.103	12.8
Netherlands	0.744	0.750	0.783	0.039	5.3
Sweden	0.855	0.804	0.845	−0.010	−1.2
Average	0.784	0.793	0.832	0.048	6.2

Notes: Actual years covered are: Australia, Finland, Sweden, United Kingdom and United States: 1980, 1990, 2000; France: 1980, 1990, 1998; W. Germany: 1984, 1990, 1998; Japan: 1980, 1990, 1999; Netherlands: 1985, 1990, 1999; New Zealand: 1984, 1990, 1997
All earnings are gross of taxes, except France which reports net earnings
Source: OECD Earnings Database

annual earnings of full-time, year-round workers (Panel B). As may be seen in the table, women uniformly have lower wage rates than men. However, in all but one case, the gap fell between 1980 and 2000. (The exception is Sweden where the pay gap rose by one percentage point from an already low level in 1980.) For example, between 1980 and 2000, the ratio of women's to men's

wages rose from an average of 69 to 77 per cent among the countries in Panel A, and from an average of 78 to 83 per cent among countries in Panel B. (US Current Population Survey data indicate that the gender gap in annual pay for full-time, year-round workers is slightly higher than the gender gap in weekly earnings for full-time workers, suggesting that the gap in weekly wages for the countries in Panel B may be even smaller than the figures shown in Table 2.) The gender wage gap is especially small in France, Australia, Sweden and New Zealand. The United States had one of the larger gaps in 1980, with only Japan having a lower female to male wage ratio. Over the 1980–2000 period, the gender wage gap fell by more, both absolutely and relatively, in the United States than in any of the other countries shown. Nonetheless, even in 2000 the gender wage gap remained relatively high in the United States, as other wage differentials such as those for education, cognitive ability or union membership have historically been (Blau and Kahn 2002, 2005). As we shall see, the pattern of international differences in the gender wage gap is useful in shedding light on the impact of labour market institutions on the gender gap. This is because there are very large differences in the types of such institutions across the economically advanced countries that have consequences for the size of the gender wage gap.

Explanations for the Gender Wage Gap

Traditionally, economic analyses of the gender wage gap have focused on what might be termed gender-specific factors, that is, (a) gender differences in qualifications and (b) differences in the labour market treatment of men and women (or labour market discrimination). More recently, following the work of Juhn et al. (1991) on trends in race differentials, some advances have been made by considering the gender wage gap and other demographic wage differentials in the context of the overall structure of wages. Wage structure is the array of prices determined for labour market skills and the rewards to employment in particular sectors. In addition, gender-specific factors and wage structure can interact to affect the gender wage gap.

Gender-Specific Factors

Gender differences in qualifications have primarily been analysed within the human capital model (Mincer and Polachek 1974). Given the traditional division of labour by gender in the family, women tend to accumulate less labour market experience than men. Further, anticipating shorter and more discontinuous work lives, women have lower incentives to invest in market-oriented formal education and on-the-job training. Their resulting smaller human capital investments will lower their earnings relative to those of men. Working in a similar direction is Becker's (1985) model in which the longer hours women spend on housework lower the effort they put into their market jobs compared with men and hence reduce their wages.

Gender differences in occupations are also expected to result if women choose occupations for which on-the-job training is less important. Women may especially avoid jobs requiring large investments in firm-specific skills (that is, skills which are unique to a particular enterprise), because the returns to such investments are reaped only as long as one remains with a particular employer. At the same time, employers may also be reluctant to hire women for such jobs because they bear some of the costs of firm-specific training (see the discussion of statistical discrimination below).

To the extent that gender differences in outcomes are not fully accounted for by productivity differences derived from these and other sources or by compensating differences in non-wage job characteristics such as risk of injury, models of labour market discrimination offer an explanation. In Becker's (1957) model, discrimination is due to the discriminatory tastes of employers, co-workers or customers. Alternatively, in models of statistical discrimination (for example, Aigner and Cain 1977), which assume a world of uncertainty, differences in the treatment of men and women arise from differences between the two groups in employer perceptions of the expected value of productivity or in the reliability with

which productivity may be predicted. In either case, gender differences in wages or occupations may result. Another aspect of interest is the relationship between occupational segregation and a discriminatory wage gap formulated in Bergmann's (1974) overcrowding model. She argues that discriminatory exclusion of women from 'male' jobs results in an excess supply of labour in 'female' occupations, depressing wages there for otherwise equally productive workers. The same wage outcomes could also be observed if women voluntarily exclude themselves from male jobs due to gender differences in preferences for the jobs themselves or for various attributes of the jobs (for example, long hours or necessity for travel).

Two recently proposed models of discrimination suggest alternative motivations for male employees to discriminate against female coworkers than the personal prejudices assumed in the Becker model, particularly for resisting the introduction of women into traditionally male occupations. In Akerlof and Kranton (2000), occupations are associated with societal notions of 'male' and 'female', leading men to resist the entry of women due to the loss in male identity (or sense of self) that this would entail. In Goldin (2002), the entry of women is viewed as reducing the prestige of the occupation, based on perceptions that women are, on average, less productive.

An additional gender-specific factor potentially contributing to the observed gender wage gap is labour force selectivity. While one would ideally like to have evidence on the potential wage offers available to each individual in the population, we typically observe wages only for those who are actually employed. If there are unobserved differences in skills or labour market prospects between the non-employed and the employed, focusing on measured wages may give a misleading picture of the wage offers received by women relative to men (Heckman 1979), both at a point in time and for trends over time and differences across countries.

Wage Structure

The human capital model suggests that men and women tend to have different levels of labour market qualifications (especially work experience) and to be employed in different occupations and perhaps in different industries. Discrimination models too suggest that women may be segregated into different sectors of the labour market. This implies that the overall returns to skills and the size of premia for employment in particular sectors potentially play an important role in determining the gender wage gap. All else equal, the larger the returns to skills and the larger the rents received by individuals in predominantly-male sectors, the larger will be the gender wage gap. The framework provided by wage structure is particularly useful in analysing changes over time in gender differentials or differences across countries in gender gaps.

Interactions Between Gender-Specific Factors and Wage Structure

While gender specific factors and wage structure each potentially play a distinct role in affecting the gender wage gap, they are likely to interact, making it sometimes difficult to disentangle their separate effects. For example, as discussed in more detail below, since the 1970s, the labour market returns to skills such as education, specialized training and experience have risen in many countries, likely due in part to technological change, including computerization. To the extent that the prices of skills for which women have a relative deficit have risen, such changes in wage structure will raise the gender wage gap. However, technological change itself is not likely to have gender neutral effects on labour demand, given occupational and industrial segregation patterns by gender. So, for example, it is likely that computerization has reduced the demand for blue-collar production labour and therefore lowered the relative demand for sectors where men are disproportionately represented (Weinberg 2000; Autor et al. 2003).

These types of analyses raise the question of what the appropriate measure of wage structure or labour market prices is. For a number of reasons it is often viewed as appropriate to use male prices/returns as the measure of overall wage structure, with the maintained hypothesis that they therefore affect women's relative wages as well. For one

thing, it is believed that men do not encounter discrimination and thus their returns are not contaminated by discrimination, although it is acknowledged that, were gender discrimination to be eliminated, male as well as female prices would likely change (for example, the supply of labour to some traditionally male occupations might increase). For another, the estimate of male prices is less likely than female prices to be influenced by selection bias or workforce interruptions. However, some research suggests that the connection between the male wage structure and women's labour market outcomes may be complicated. For example, Fortin and Lemieux (1998) present a model in which there is a fixed hierarchy of jobs. As women move up the hierarchy, they replace some men who previously would have had middle-level positions, bumping them down the hierarchy. Thus, increases in women's human capital or reductions in employment discrimination against women may cause increases in male wage inequality, and the male wage distribution may change even with no changes in the overall wage distribution. Topel (1994) makes a related argument to the effect that high-skill women compete with low-skill men in the labour market and, thus, that increases in the supply of high-skill women directly lower the real wages of low-skill men (through this increase in supply) and thereby raise male wage inequality.

Evidence on Human Capital, Discrimination and the Gender Wage Gap

The typical approach to analysing the sources of the gender wage gap is to estimate wage regressions specifying the relationship between wages and productivity-related characteristics for men and women. While it would be preferable to analyse total compensation (including non-wage benefits and compensating differentials for job amenities), virtually all studies focus on money wages since data on total compensation are generally not available. The gender wage gap may then be statistically decomposed into two components: one due to gender differences in measured characteristics, and another which is 'unexplained' and potentially due to discrimination. Such empirical studies provide evidence consistent with both human capital differences and labour market discrimination in explaining the gender wage gap.

One problem with this approach is that evidence for discrimination relies on the existence of a residual gender wage gap, which cannot be explained by gender differences in measured qualifications. This accords well with the definition of labour market discrimination, that is, pay differences between groups that are not explained by productivity differences, but these may also reflect group differences in unmeasured qualifications. If men are more highly endowed with respect to these omitted variables then we would overestimate discrimination. And, conversely, to the extent that women are more highly endowed with respect to the omitted variables, discrimination would be underestimated. Another case in which discrimination would be underestimated would be if some of the factors controlled for (for example, occupation or tenure with the employer) themselves reflect the impact of discrimination.

Another challenge to empirically decomposing the gender wage gap into its constituent parts is the existence of feedback effects. The traditional division of labour in the family may influence women's market outcomes through its effects on their acquisition of human capital and on rationales for employer discrimination against them. But it is also the case that, by lowering the market rewards to women's human capital investments and labour force attachment, discrimination may reinforce the traditional division of labour in the family (for example, Weiss and Gronau 1981). Even small initial discriminatory differences in wages may cumulate to large ones as men and women make human capital investment and time allocation decisions on the basis of them.

Representative Findings from Statistical Analyses

Representative findings from analyses of this type may be illustrated by results from three recent

studies of the gender wage gap in the United States (Blau and Kahn 2006), Denmark (Datta Gupta et al. 2006), and Sweden (Edin and Richardson 2002). Each of these studies uses databases that have information on actual labour market experience, a variable that is crucial for the analysis and is often not available in nationally representative data sets.

For the United States, Blau and Kahn (2006) found a female–male ratio for average hourly earnings of 79.7 per cent in 1998. In light of the issues discussed above, they considered results when only human capital variables (that is, education and labour market experience) and race were taken into account, and results additionally controlling for occupation, industry and unionism. While gender differences in educational attainment were fairly small, and actually favoured women, men had more full-time work experience than women. Controlling for human capital, women earned 81 per cent of what men earned; the relatively small increase in the human-capital adjusted ratio compared with the raw ratio reflects the offsetting effects of adjusting for gender differences in education and experience. The gender ratio rose to 91 per cent when industry, occupation and union status were additionally controlled for. For Denmark, Datta Gupta et al. (2006) found an unadjusted gender ratio of 81.1 per cent in 1995. This rose to 83.2 per cent controlling for schooling and experience, and to 86.2 per cent additionally controlling for industry, occupation and region. Edin and Richardson (2002) found qualitatively similar results for Sweden in 1991. Thus, in all three countries measured characteristics explained some but not all of the gender wage gap.

Studies such as those discussed above suggest that gender differences in human capital (especially experience) can be an important factor helping to account for the gender wage gap at any given point in time. In the United States, improvements in women's relative experience were an important factor in explaining the rise in women's relative wages during the 1980s (Blau and Kahn 1997, 2006; O'Neill and Polachek 1993), while increases in women's relative experience and education both contributed to female wage gains in the 1990s (Blau and Kahn 2006). And since 1980 the unexplained wage gap in the United States has fallen, a finding consistent with a decline in discrimination or improvements in women's unmeasured characteristics, and also, as we shall see below, with shifts in relative demand favouring women. Sample selectivity can also affect measured gender wage gaps. For example, using different methodologies, Blau and Kahn (2006) and Mulligan and Rubinstein (2005) both find a role for selectivity in explaining these wage trends.

Additional Evidence on Discrimination

A problem with the types of statistical analyses just discussed is that evidence of discrimination is based on a residual or unexplained gender wage gap that is susceptible to a variety of interpretations, of which labour market discrimination is only one. Two lines of empirical research on discrimination pursue alternative approaches which lend additional support to the finding of discrimination.

First are two studies that use an experimental approach. Neumark (1996) analysed the results of a hiring 'audit' in which male and female pseudo-job seekers were given similar résumés and sent to apply for waiter or waitress jobs at the same set of Philadelphia restaurants. In high-priced restaurants where earnings of workers are generally higher than in the other establishments, a female applicant's probability of getting an interview was 40 percentage points lower than a male's, and her probability of getting an offer was 50 percentage points lower. A second study, by Goldin and Rouse (2000), examined the impact of the 'natural experiment' in which major symphony orchestras in the United States adopted 'blind' auditions. In a blind audition, a screen is used to conceal the identity of the candidate. Using data from actual auditions, the authors found that the screen substantially increased the probability that a woman would advance out of preliminary rounds and be the winner in the final round. Goldin and Rouse (2000) used their parameter estimates to conclude that the switch to blind auditions can explain one

quarter of the increase in the female percentage in the top five symphony orchestras in the United States from less than 5 per cent of all players in 1970 to 25 per cent in 1996.

A second source of additional evidence on discrimination is provided by studies that examine predictions of Becker's (1957) discrimination model and obtain results which are consistent with the model and hence with discrimination against women. Becker and others have pointed out that competitive forces should reduce or eliminate employer discrimination in the long run because the least discriminatory firms, which hire more lower-priced female labour, would have lower costs of production and should drive the more discriminatory firms out of business. For this reason, Becker suggested that discrimination would be more severe in firms or sectors that are shielded to some extent from competitive pressures. Consistent with this reasoning, Hellerstein et al. (2002) found that, among plants with high levels of product market power, those employing relatively more women were more profitable. Similarly, Black and Strahan (2001) report that, with the deregulation of the banking industry beginning in the mid-1970s, the gender wage gap in banking declined. And Black and Brainerd (2004) found that increasing vulnerability to international trade reduced apparent gender wage discrimination in concentrated industries, again as predicted by the Becker model.

Possible Sources of the Unexplained Gender Wage Gap

While there appears to be evidence from a variety of approaches that is consistent with discrimination against women in the labour market, this does not mean that the full unexplained gap estimated in traditional approaches may be attributed to discrimination. Some of the residual gap may be due to the impact of childbearing on women's wages. This is not a factor that can be examined simply by including a control for number of children in a wage regression, since the coefficient on children variables may be influenced by self-selection into motherhood and the endogeneity of number of children. Research that addresses some of these issues suggests a negative effect of children on wages, even when labour market experience is controlled for (for example, Waldfogel 1998). This may reflect the fact that, in the past, the birth of a child often meant that a woman withdrew from the labour force entirely, breaking her tie to her employer and forgoing the returns to any firm-specific training she might have acquired, as well as any rewards for having made an especially good job match.

Another possible source of the unexplained wage gap is noncognitive skills/traits or what Fortin (2005) terms 'soft factors'. For example, experimental evidence suggests there are gender differences in competitiveness (for example, Gneezy et al. 2003), and negotiating skills (Babcock and Laschever 2003). Fortin (2005) examines the impact of a number of noncognitive traits and attitudes in a wage regression context. While such findings are informative in elucidating the omitted factors that lie behind the unexplained gap in traditional wage equations, as Fortin acknowledges, the coefficients on soft factors in a wage equation cannot necessarily be given a causal interpretation. Both wages and attitudes, for example, may be determined by the same exogenous factor. And, as in the case of the traditional productivity proxies discussed above, there may be important feedback effects from differential treatment in the labour market to noncognitive traits. So, for example, income expectations may influence wages through negotiating behaviour or effort, but the source of women's lower income expectations could be, at least in part, anticipation of labour market discrimination. Nor is it clear that all such omitted factors favour men. Borghans et al. (2005) argue for a female advantage in interpersonal interactions, which they proxy by altruism.

Just as the importance of gender differences in the traditional human capital variables may change over time, thus helping to account for the decline in the gender wage gap, so may the impact of noncognitive traits. In this regard, it is interesting that Fortin (2005) finds evidence that gender differences in work attitudes were much smaller in 2000 than in 1986. Further, Borghans et al. (2005)

find evidence of a growing importance of interpersonal interactions (in part due to increased computer use) in affecting wages that can help explain rapidly rising female relative wages in the 1980s as well as a slower rate of increase in the 1990s.

The Impact of Policy

Women's relative skills and the degree of employer discrimination can be affected by government policies directed at issues of combining work and family as well as equal employment opportunity laws. For example, many countries have enacted paid parental leave mandates which give parents who take time off to care for children or other relatives an entitlement to their jobs upon returning from the leave. While such policies may encourage firm-specific investments, thus raising women's relative wages (since parental leave is much more likely to be taken by women than men), they may also encourage labour force withdrawal for longer periods of time than otherwise, reducing women's accumulation of experience. Mandated paid leaves, particularly of long duration, may also diminish women's opportunities by increasing employer costs of hiring women and hence providing incentives to discriminate against them. Thus, the effect of parental leaves on the gender wage gap is theoretically ambiguous. Ruhm (1998) in fact finds in a study of 16 Western industrialized countries that, other things equal, short mandated paid parental leaves lead to higher relative wages for women, while longer leaves lead to a higher gender wage gap. These results suggest that a number of offsetting factors may be at work, with the positive impact dominating for short leaves and the negative effect dominating for long periods of mandated parental leave.

While virtually all industrialized countries have enacted legislation outlawing employment discrimination against women, in some countries government intervention is more dramatic than in others. The major approach in the United States involves enforcement of antidiscrimination legislation, including equal employment opportunity as well as equal pay for equal work. Further, under some circumstances, affirmative action, or 'proactive steps … to erase differences between women and men, minorities and nonminorities, etc.' (Holzer and Neumark 2000, p. 484), is also required or voluntarily adopted by employers. There is some evidence for the United States of a positive effect of government anti-discrimination policies on women's earnings and occupations. Studies focusing specifically on the impact of affirmative action also suggest modest employment and wage gains for women attributable to this programme. (For a summary, see Blau et al. 2006, pp. 240–245.)

'Comparable worth' or equal pay for work of equal value (that is, even if men and women are doing different jobs) constitutes a stronger form of government intervention. In evaluating the impact of such a policy it is interesting to look at studies focusing on Australia, which has adopted government mandates in this area nationwide, and the United States, where such policies are limited to selected state or local government employees (Gregory and Duncan 1981; Killingsworth 1990; O'Neill et al. 1989). One would expect that if such policies lower the gender wage gap, they might also lead to a decrease in women's relative employment due to employer demand effects. Gregory and Duncan (1981) in fact find such a pattern: the gender wage gap fell dramatically immediately after the Australian tribunal began implementing comparable worth policies in the early 1970s, but female employment grew less rapidly than one would have predicted in the absence of the wage intervention. Similarly, in studies of the impact of the comparable worth policies in state governments in the United States, small positive wage and negative employment effects for women have been found (Killingsworth 1990; O'Neill et al. 1989).

While these results for the impact of comparable worth in Australia and the United States on women's employment are consistent with the existence of competitive labour markets, to the extent that the labour market is characterized by monopsony, government-mandated wage increases for women need not result in a reduction in women's employment levels. Manning (1996) interprets the impact of the UK Equal Pay Act of 1970 and the Sex Discrimination Act of 1975 in

this light. Specifically, he shows that these laws led to a major reduction in the gender wage gap in the United Kingdom with no apparent employment losses for women: after the legislation, wage changes and employment changes within industries were strongly positively related for women but much less so for men.

Differential monopsony power facing men and women could help to explain the existence of the gender wage gap (Madden 1973) in general, as well as Manning's (1996) results for the policy intervention. For this explanation of the wage gap to make sense, women's supply of labour to the *firm* must be less wage elastic than men's, giving employers greater monopsony power over women than men. This might seem counter-intuitive at first, in that there is clear evidence that women have a *larger* own-wage elasticity of labour supply to the *labour market* than men, although in the United States the gender difference has been decreasing as women's elasticity has declined since 1980 (Blau and Kahn 2007; Heim 2007). However, a variety of factors could still potentially result in women having a smaller responsiveness to wage changes at the firm level. Perhaps the most intriguing possibility is discrimination itself. Black (1995) develops a model in which search costs give employers a degree of monopsony power. If there is discrimination against women, women will face higher search costs than men, increasing employers' monopsony power over them.

Evidence on gender differences in labour supply at the firm level is mixed. On the one hand, Viscusi (1980), Blau and Kahn (1981) and Light and Ureta (1992) all find that, for the United States, women's quit rates are at least as wage responsive as men's, suggesting that the monopsony model may have limited application in the United States. On the other hand, Barth and Dale-Olsen (1999) found that men's turnover in Norway is more wage-elastic than women's. Thus, Norwegian employers could potentially exercise differential monopsony power over women. Of course, the degree to which Norway's centralized wage-setting system would allow this to take place is an empirical question (Kahn 1998).

Evidence on the Impact of Wage Structure on the Gender Wage Gap

The impact of wage structure on the gender wage gap is best studied in a comparative context. Since wage structure may differ across countries and change over time, investigations of the impact of wage structure have focused on (a) international differences in the gender wage gap at a specific point in time, and (b) changes in the gender wage gap in one country over time. A useful framework for analysing the impact of wage structure on demographic wage differentials was devised by Juhn, Murphy and Pierce (JMP) (1991) in their analysis of changes in black workers' relative wages in the United States. Blau and Kahn (for example, 1992, 1996b) have adapted their framework to studying international differences in the gender wage gap as follows.

Suppose we have for male worker i in country j the following wage equation:

$$Y_{ij} = B_j' X_{ij} + e_{ij} = B_j' X_{ij} + F_j^{-1}(\theta_{ij}), \quad (1)$$

where Y is log of wages, B is a coefficient vector, X is a vector of productivity-related characteristics, e is a disturbance term, $F^{-1}(\cdot)$ is the inverse cumulative distribution function of male log wage residuals, and θ is individual i's percentile in the male residual distribution. Estimating Eq. (1) separately for each of two countries, differences in the gender wage gap may be decomposed into components due to inter-country differences in: (a) gender differences in the X variables; (b) the male wage coefficients B; (c) women's position in the male residual distribution (θ); and (d) the residual distribution $F(\cdot)$. Components (a) and (c) represent gender-specific factors: inter-country differences in women's relative measured productivity (component a) and in women's placement in the distribution of male wage residuals (component c). The latter can represent discrimination or unmeasured productivity differences. Components (b) and (d) represent the potential effects of wage structure: measured prices (component b) and the prices of unmeasured skills or rents due to unmeasured representation in favourable sectors (component d). Note that the

sum of components (c) and (d) corresponds to the unexplained gap in a traditional decomposition of the gender wage gap, and component (c) may be viewed as the unexplained gap adjusted for differences in unmeasured prices.

Some examples may help to illustrate these components. It is straightforward that, if one country has a larger gender gap in experience, it will have a larger gender wage gap. But it is less obvious in the absence of this decomposition that, if the return to experience, which is part of the B vector, is higher in one country than another, then this difference will contribute to a higher gender wage gap in the first country, since women on average have less experience than men. Or suppose that X does not include data on the specific firm in which a worker is employed. If a country has especially high inter-firm wage differentials (part of the residual wage distribution $F(\ )$) and if women are employed in low-wage firms on average, then this unmeasured price effect will raise that country's gender wage gap. The same decomposition can also be used to explain changes in the gender wage gap over time within a country.

It may be noted that the interpretation of the residual proposed by JMP has been questioned by Suen (1997). For further discussion of this issue, see our discussion below on the assumption in the JMP decomposition that male prices and male residuals are relevant indicators of the prices facing women in the labour market. A fuller discussion is provided in Blau and Kahn (2003).

The JMP decomposition has been used by Blau and Kahn (1992, 1996b) and Kidd and Shannon (1996) to study international differences in the gender wage gap at a point in time. For example, Blau and Kahn (1996b) compared the US gender wage gap in the late 1980s with that in nine other countries (Australia, Austria, West Germany, Hungary, Italy, Norway, Sweden, Switzerland and the United Kingdom). They found that, on average, the ratio of women's to men's pay was 4.3 percentage points lower in the United States than in the other countries: 65.4 per cent as against 69.7 per cent. US women had better measured characteristics and were placed higher in the distribution of male residuals than women in the other countries, suggesting that gender-specific factors could not explain the higher US gender wage gap. However, measured and unmeasured prices together had large effects raising women's relative wages in the other countries compared with the United States. Wage structure was thus sufficient to explain more than the full amount of the difference between the US gender wage gap and that in other countries. Blau and Kahn (1992, 1996b) interpreted this pattern as reflecting the impact of international differences in labour market institutions. In the other countries, unions cover a much larger portion of the labour market than in the United States, and wage-setting is much more centralized. Centralized collective bargaining tends to reduce wage differentials through the negotiation of relatively high wage floors, which would raise the relative wages of anyone near the bottom of the distribution, including women (Blau and Kahn 1996a).

Kidd and Shannon (1996) also found an important role for wage structure in their study of the gender wage gaps in Australia and Canada for 1989–1990. Specifically, the gender gap in hourly wages was about 0.14 log points lower in Australia than in Canada. They found that 0.05–0.09 log points of this difference was due to the combined effects of observed and unobserved prices. This result is similar to that in Blau and Kahn (1992, 1996b) in that Australia has much higher coverage by collective bargaining than Canada.

The JMP decomposition has also been used to study the impact of wage structure on changes in the gender wage gap over time within a country. For example, in Sweden between 1968 and 1974, the trade union movement engineered a major compression of wages. Edin and Richardson (2002) used the JMP decomposition to find that wage structure, especially unobserved prices, contributed to a reduction of the gender wage gap during this period. Moreover, Datta Gupta et al. (2006) used a version of the JMP decomposition to study changes in the Danish gender wage gap between 1983 and 1995. This was a period of increased decentralization of the wage determination process, a development that is expected to lead to rising labour market prices and therefore a rising gender wage gap. The authors indeed found

that the gender wage gap in Denmark increased during this period and that most of the increase can be accounted for by rising unmeasured prices.

Finally, this approach has been applied to understanding the trends in the gender wage gap in the United States. Blau and Kahn (1997) used the JMP decomposition to study the apparent paradox of a substantial decrease in the gender wage gap in the United States during the 1979–1988 period, a time of rising skill prices. They found that women were able to overcome the negative effects of these price changes by improving their measured human capital and by moving up the distribution of male residuals. The authors further noted that the process leading to higher skill prices in the United States might not have been gender-neutral. Specifically, it is likely that part of the explanation for this development involves skill-biased technical change in which the demand for white-collar labour rose relative to blue-collar labour, a change that, given gender differences in occupational distributions, in effect raises the demand for women workers. Thus, while skill prices were rising, contributing to a reduction in women's relative wages, developments such as computerization and perhaps outsourcing of production labour disproportionately lowered the demand for male labour (Welch 2000; Weinberg 2000; Autor et al. 2003). In the context of the JMP decomposition, such changes in labour demand would be reflected in higher placement of women in the distribution of male residuals (leading to a decrease in the conventional unexplained gender wage gap). As noted previously, convergence in the gender wage gap in the United States slowed during the 1990s. Using the JMP decomposition, Blau and Kahn (2006) found that the major reason for the slowdown was the considerably smaller narrowing of the unexplained gender wage gap in the 1990s than in the 1980s. This raises the possibility that the types of demand shifts favouring women that we have outlined here were smaller in the 1990s than in the 1980s, and Blau and Kahn present some evidence that is consistent with this as at least a partial explanation for the smaller decrease in the unexplained gender gap in the latter period.

The JMP decomposition assumes that male prices and male residuals are relevant indicators of the prices facing women in the labour market. Some support for this assumption is provided by the fact that wage coefficients and residual distributions have changed similarly for men and women over time in the United States and are similar to each other within countries at a point in time (Blau and Kahn 2002). But it is possible to directly test whether male wage compression leads to a smaller gender wage gap, and Blau and Kahn (2003) have done so by compiling a microdata-set for 22 countries over the 1985–1994 period. They find, looking across countries, that the gender wage gap is positively affected by a country's male skill prices (that is, the level of male wage inequality adjusted for measured characteristics), as well as by the relative net supply of women (that is, supply net of demand). A likely interpretation is that more compressed male wages are an indicator of smaller wage differentials in general, as suggested above in our discussion of centralized wage-setting institutions. Bolstering this interpretation is the authors' further finding that, other things equal, greater coverage by collective bargaining reduces the gender wage gap. It thus appears that high wage floors negotiated by unions serve to lower the gender wage gap.

If labour markets are competitive, then union-negotiated wage floors should lower female relative employment. And this is precisely what Bertola et al. (2007) find in a study of relative employment in 17 countries over the 1960–1996 period. Specifically, they find that greater coverage by highly centralized unions lowers female employment and raises female unemployment compared with men's. This suggests that unionization can raise women's relative wages at the expense of lowering their employment. This in turn suggests that Manning's (1996) evidence in favour of monopsony in the United Kingdom may describe an exceptional case in the OECD.

Future Prospects

While it is difficult to speculate about the future, Tables 1 and 2 do suggest some convergence across countries in both female labour force

participation (absolutely and relative to men) and in the gender wage gap. For example, calculations based on the data in Tables 1 and 2 indicate that the standard deviation across countries in the ratio of women's to men's labour force participation rates fell from 0.128 in 1980 to 0.072 in 2000, and that the standard deviation of the gender wage ratio fell steadily from 0.086 in 1980 to 0.066 by 2000. While the gender wage ratio appears to be converging at around 80 per cent for several of the countries in Table 1, further changes are not precluded. Throughout the OECD, women's education has been rising relative to men's, a trend that shows no sign of ending (Goldin et al. 2006). Technological change, which has likely raised women's relative wages through demand effects, will probably continue and could even accelerate. Going against these trends is the likely continued decentralization of wage-setting institutions in many Western countries, spurred in part by globalization (Katz 1993).

See Also

- ▶ Affirmative Action
- ▶ Family Decision Making
- ▶ Feminist Economics
- ▶ Gender Differences (Experimental Evidence)
- ▶ Gender Roles and Division of Labour
- ▶ Inequality (International Evidence)
- ▶ Labour Supply

Bibliography

Aigner, D., and G. Cain. 1977. Statistical theories of discrimination in labor markets. *Industrial & Labor Relations Review* 30: 175–187.

Akerlof, G.A., and R.E. Kranton. 2000. Economics and identity. *Quarterly Journal of Economics* 115: 715–753.

Autor, D.H., F. Levy, and R.J. Murnane. 2003. The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics* 118: 1279–1333.

Babcock, L., and S. Laschever. 2003. *Women don't ask: Negotiation and the gender divide*. Princeton/Oxford: Princeton University Press.

Barth, E., and H. Dale-Olsen. 1999. Monopsonistic discrimination and the gender wage gap. Working paper no. 7197. Cambridge, MA: NBER.

Becker, G.S. 1957. *The economics of discrimination*. Chicago: University of Chicago Press.

Becker, G.S. 1985. Human capital, effort, and the sexual division labor. *Journal of Labor Economics* 3(Supp.): S33–S58.

Bergmann, B. 1974. Occupational segregation, wages, and profits when employers discriminate by race or sex. *Eastern Economic Journal* 1: 103–110.

Bertola, G., F.D. Blau, and L.M. Kahn. 2007. Labor market institutions and demographic employment patterns. *Journal of Population Economics* 20: 833–867.

Black, D.A. 1995. Discrimination in an equilibrium search model. *Journal of Labor Economics* 13: 309–334.

Black, S.E., and E. Brainerd. 2004. The impact of globalization on gender discrimination. *Industrial & Labor Relations Review* 57: 540–559.

Black, S.E., and P.E. Strahan. 2001. The division of spoils: Rent-sharing and discrimination in a regulated industry. *American Economic Review* 91: 814–831.

Blau, F.D., and L.M. Kahn. 1981. Race and sex differences in quits by young workers. *Industrial & Labor Relations Review* 34: 563–577.

Blau, F.D., and L.M. Kahn. 1992. The gender earnings gap: Learning from international comparisons. *American Economic Review* 82: 533–538.

Blau, F.D., and L.M. Kahn. 1996a. International differences in male wage inequality: Institutions versus market forces. *Journal of Political Economy* 104: 791–837.

Blau, F.D., and L.M. Kahn. 1996b. Wage structure and gender earnings differentials: An international comparison. *Economica* 63(Supp.): S29–S62.

Blau, F.D., and L.M. Kahn. 1997. Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics* 15: 1–42.

Blau, F.D., and L.M. Kahn. 2002. *At home and abroad: U.S. labor market performance in international perspective*. New York: Russell Sage.

Blau, F.D., and L.M. Kahn. 2003. Understanding international differences in the gender pay gap. *Journal of Labor Economics* 21: 106–144.

Blau, F.D., and L.M. Kahn. 2005. Do cognitive test scores explain higher US wage inequality? *Review of Economics and Statistics* 87: 184–193.

Blau, F.D., and L.M. Kahn. 2006. The US gender pay gap in the 1990s: Slowing convergence. *Industrial & Labor Relations Review* 60: 45–66.

Blau, F.D., and L.M. Kahn. 2007. Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor Economics* 25: 393–438.

Blau, F.D., M.A. Ferber, and A.E. Winkler. 2006. *The economics of women, men, and work*. 5th ed. Englewood Cliffs: Prentice-Hall.

Borghans, L., B. ter Weel, and B.A. Weinberg. 2005. People people: Social capital and the labor-market outcomes of underrepresented groups. Discussion paper no. 1494, Institute for the Study of Labor (IZA).

Datta Gupta, N., R.L. Oaxaca, and N. Smith. 2006. Swimming upstream, floating downstream: Comparing women's relative wage position in the US and Denmark. *Industrial & Labor Relations Review* 59: 243–266.

Edin, P.-A., and K. Richardson. 2002. Swimming with the tide: Solidarity wage policy and the gender earnings gap. *Scandinavian Journal of Economics* 101: 49–67.
Fortin, N. 2005. Greed, altruism, and the gender wage gap. Working paper, University of British Columbia.
Fortin, N.M., and T. Lemieux. 1998. Rank regressions, wage distributions, and the gender gap. *Journal of Human Resources* 33: 610–643.
Gneezy, U., M. Niederle, and A. Rustichini. 2003. Performance in competitive environments: Gender differences. *Quarterly Journal of Economics* 118: 1049–1074.
Goldin, C. 2002. A pollution theory of discrimination: Male and female differences in occupations and earnings. Working paper no. 8985. Cambridge, MA: NBER.
Goldin, C., and C. Rouse. 2000. Orchestrating impartiality: The impact of 'blind' auditions on female musicians. *American Economic Review* 90: 715–741.
Goldin, C., L. Katz, and I. Kuziemko. 2006. The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives* 20: 133–156.
Gregory, R.G., and R.C. Duncan. 1981. Segmented labor market theories and the Australian experience of equal pay for women. *Journal of Post Keynesian Economics* 3: 403–428.
Heckman, J.J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.
Heim, B.T. 2007. The incredible shrinking elasticities: Married female labor supply, 1979–2003. *Journal of Human Resources* (forthcoming).
Hellerstein, J.K., D. Neumark, and K. Troske. 2002. Market forces and sex discrimination. *Journal of Human Resources* 37: 353–380.
Holzer, H., and D. Neumark. 2000. Assessing affirmative action. *Journal of Economic Literature* 38: 483–568.
Juhn, C., K.M. Murphy, and B. Pierce. 1991. Accounting for the slowdown in black-white wage convergence. In *Workers and their wages*, ed. M. Kosters. Washington, DC: AEI Press.
Kahn, L.M. 1998. Against the wind: Bargaining recentralisation and wage inequality in Norway, 1987–1991. *Economic Journal* 108: 603–645.
Katz, H.C. 1993. The decentralization of collective bargaining: A literature review and comparative analysis. *Industrial & Labor Relations Review* 47: 3–22.
Kidd, M.P., and M. Shannon. 1996. The gender wage gap: A comparison of Australia and Canada. *Industrial & Labor Relations Review* 49: 729–746.
Killingsworth, M. 1990. *The economics of comparable worth*. Kalamazoo: W.E. Upjohn Institute for Employment Research.
Light, A., and M. Ureta. 1992. Panel estimates of male and female job turnover behavior: Can female nonquitters be identified? *Journal of Labor Economics* 10: 156–181.
Madden, J.F. 1973. *The economics of sex discrimination*. Lexington: DC Heath.
Manning, A. 1996. The equal pay act as an experiment to test theories of the labour market. *Economica* 63: 191–212.
Mincer, J., and S. Polachek. 1974. Family investments in human capital: Earnings of women. *Journal of Political Economy* 82(Part 2): S76–S108.
Mulligan, C.B., and Y. Rubinstein. 2005. Selection, investment, and women's relative wages since 1975. Working paper no. 11159. Cambridge, MA: NBER.
Neumark, D.M. 1996. Sex discrimination in restaurant hiring: An audit study. *Quarterly Journal of Economics* 111: 915–941.
O'Neill, J., and S. Polachek. 1993. Why the gender gap in wages narrowed in the 1980s. *Journal of Labor Economics* 11: 205–228.
O'Neill, J., M. Brien, and J. Cunningham. 1989. Effects of comparable worth policy: Evidence from Washington State. *American Economic Review* 79: 305–309.
OECD. 2004. *Employment outlook*. Paris: OECD.
OECD (Organisation for Economic Co-operation and Development). 1990. *Employment outlook*. Paris: OECD.
Ruhm, C.J. 1998. The economic consequences of parental leave mandates: Lessons from Europe. *Quarterly Journal of Economics* 113: 285–317.
Suen, W. 1997. Decomposing wage residuals: Unmeasured skill or statistical artifact? *Journal of Labor Economics* 15: 555–566.
Topel, R.H. 1994. Regional labor markets and the determinants of wage inequality. *American Economic Review* 84: 17–22.
Viscusi, W.K. 1980. Sex differences in worker quitting. *Review of Economics and Statistics* 62: 388–398.
Waldfogel, J. 1998. The family gap for young women in the United States and Britain: Can maternity leave make a difference? *Journal of Labor Economics* 16: 505–545.
Weinberg, B. 2000. Computer use and the demand for female workers. *Industrial & Labor Relations Review* 53: 290–308.
Weiss, Y., and R. Gronau. 1981. Expected interruptions in labour force participation and sex-related differences in earnings growth. *Review of Economic Studies* 48: 607–619.
Welch, F. 2000. Growth in women's relative wages and in inequality among men: One phenomenon or two? *American Economic Review* 90: 444–449.

Wood, Stuart (1853–1914)

George J. Stigler

Keywords

Factor substitution; Marginal productivity theory; Technological substitution; Wages fund; Wood, S.

JEL Classifications
B31

A member of a prominent Philadelphia Quaker family, Wood was briefly active in economics twice in his life. The first period was 1873–5, when he received at Harvard the first Ph.D. in economics in the United States. The second period was 1888–90, when he wrote three first-class articles on wage theory. His primary interests during his adult life were in business and finance.

In the two years Wood was at Harvard he took courses in economics and its history, chiefly from Professor Charles Dunbar, and wrote an essay on 'A Review of the "Principles of Social Science" by Henry C. Carey'. It was not an impressive piece, even allowing for the time, the age (21) of the writer and the extreme vulnerability of the target.

It is all the more impressive that 13 years later he wrote two fine articles on the marginal productivity theory and one on the history of the wages-fund theory. Wood must be acknowledged to be an independent discoverer of the marginal productivity theory, an honour he shares with Marshall, Edgeworth, Barone, Wicksell, Clark and other major economists. Wood's version was not mathematical, but it synthesized two important dimensions of substitution between capital and labour: the substitution between industries with different capital–labour ratios, and the substitution within enterprises. The formulation was a skilful synthesis incorporating consumer demands and factor supplies as well as technological substitution.

Wood's final contribution was a history of the wages-fund doctrine (which was to be treated no more penetratingly by Harvard's second Ph.D. in economics, F.W. Taussig). Perhaps one should mention one other, involuntary role Wood played in the study of the history of economics: he was the victim of a thinly disguised, utterly unfounded charge of plagiarism (of Lord Lauderdale) in the *Journal of Political Economy* in 1894.

Selected Works

1888–9. A new view of the theory of wages. Pts I–II. *Quarterly Journal of Economics* 3, 60–86; 462–480.
1889. The theory of wages. *Publications of the American Economic Association* 4: 5–35.
1890. A critique of wage theories. *Annals of the American Academy of Political and Social Science* 1: 426–461.

Bibliography

Stigler, G.J. 1947. Stuart Wood and the marginal productivity theory. *Quarterly Journal of Economics* 61: 640–649.

Worker Participation and Profit Sharing

John Pencavel

Keywords
Democracy; Incentives; Productivity; Profit-sharing; Worker participation

JEL Classifications
J33; J53; J54

Market economies are called 'capitalist' because in such economies most production is carried out in organizations owned by those who supply the firms' financial capital. A firm is 'owned' by its capital investors because, first, the capital investors claim the firm's net receipts or profits and, second, they have the authority to direct and manage (often indirectly) the firm's activities.

Yet in all market economies some production takes place in firms where these two dimensions of ownership are embodied in those who supply labour rather than capital. In this instance, workers enjoy as incomes the firm's net receipts and the workers hire individuals to supervise and organize production. Capital may be obtained from the workers' savings or from loans from financial intermediaries. Examples of worker cooperatives include the plywood companies in the Pacific

Northwest of the United States, the kibbutzim in Israel, and the Mondragon group in the Basque country of Spain.

Many enterprises fall between these two limiting cases. These other firms are characterized by the owners either sharing net revenues with others – 'profit-sharing' – or sharing in the activities of management – 'worker participation'. (For recent research on the general issues, see the essays in Blair and Roe 1999, and Ichniowski et al. 2000.)

Profit-sharing occurs when those who have the right to consume all the firm's profits distribute a portion of them to others within the organization. Because most firms are owned by those who supply capital, profit-sharing usually occurs when some portion of profits is distributed among the rank-and-file workers.

With explicit profit-sharing, a clear formula is established linking profits and the pay of individuals. Profit-sharing is implicit when workers in firms that habitually enjoy higher profits are paid higher wages. Profit-sharing may take the form of deferred income, as when a portion of net receipts is placed in retirement accounts so that the firm's employees hold part of the assets of the firm in which they work.

A principal goal of these various profit-sharing arrangements is to affect incentives: by linking workers' compensation to the firm's success in making profits, the workers' interests are aligned more closely with the capital owners'. However, some economists reason that, when the firm's net earnings are divided among a large number of people and one individual's effort contributes little to total output, the incentive for a single individual to apply effort is meagre. What empirical evidence there is suggests that, with profit-sharing, workers monitor one another so that any tendency to shirk is checked.

When workers' pay is linked to profits, some automatic flexibility is imparted to a firm's payrolls so the effects of adverse shocks are communicated immediately and mechanically to the firm's costs. Some have suggested that, if profit-sharing payment schemes were widespread, recessions would be characterized by less unemployment. Kruse (1993) reviews profit-sharing.

'Worker participation' is a term embracing various arrangements by which workers are actively involved in the management of the enterprise where they work. These arrangements may include safety and health committees or panels to deal with worker grievances, or they may be more profound arrangements when workers are actively engaged in key management activities such as the organization of work and production. In Europe, works councils or workers' committees are empowered to be consulted and, sometimes, to share in determining any changes in the organization of production (also known as co-determination).

One argument in favour of worker participation is that participation begets productivity. Modern 'flexible' or 'lean' production techniques entail greater employee involvement in shop-floor decisions, greater teamwork, information-sharing between management and rank-and-file employees, and reduced task specialization. An extensive research literature quantifies the effect of greater worker participation on productivity. A general finding is that there are positive, though small, productivity benefits accompanying worker participation.

A second argument for worker participation is that it is the extension to the workplace of democratic governance in the political arena. In much the same way as citizens in political democracies have an important voice in choosing those who determine the provision of public goods in society so an enterprise's workers should have a voice in shaping their work environment when public goods are also prevalent. Worker participation is the application of the democratic principle to the workplace.

See Also

► Industrial Relations

Bibliography

Blair, M., and M. Roe (eds.). 1999. *Employees and corporate governance*. Washington, DC: Brookings Institution Press.

Ichniowski, C., D. Levine, C. Olson, and G. Strauss (eds.). 2000. *The American workplace: Skills, compensation,*

and employee involvement. Cambridge: Cambridge University Press.
Kruse, D. 1993. *Profit sharing: Does it make a difference?* Kalamazoo: W.E. Upjohn Institute for Employment Research.

World Bank

Albert Fishlow

Abstract

The World Bank, established in 1944, remained an important source of funding for developing countries generally through the early 1990s. Then an impressive increase in private flows reduced its overall significance. What remained was technical assistance on the one hand, and continued increasing credits of the International Development Association to the lowest income countries on the other. Despite criticism from both by the Right and the Left, the World Bank has survived, and has given voice to rising concerns about highly unequal distributions of income in the developing world, moving away from its earlier emphasis upon economic growth alone.

Keywords

Conditionality; Free trade; International Bank for Reconstruction and Development; International Development Association; International Finance Corporation; International Monetary Fund; Marshall Plan; Non-governmental organizations; Poverty alleviation; Privatization; Structural loans; World Bank; World Trade Organization

JEL Classifications

F33

The World Bank was founded in July 1944 as part of a new financial architecture for the post-Second World War period. At the inaugural meeting at Bretton Woods, the International Bank for Reconstruction and Development (IBRD) was proposed in order to satisfy demands in lesser-income countries for long-term capital, a market that had virtually disappeared with the Great Depression. The International Monetary Fund (IMF), much more a subject of debate at the time, was to satisfy anticipated short-term balance of trade needs, thereby avoiding competitive currency devaluations. A third component, the later proposed, but unfounded, International Trade Organization, was to resurrect freer trade. GATT, now the World Trade Organization (WTO), took on that function in 1947.

The Bank began in 1946 by focusing on reconstruction, which soon was successfully taken on by the Marshall Plan, and then increasingly committed itself to the multiple and changing problems of economic development within the world's poorer countries. This conversion came about much more rapidly than John Maynard Keynes, and most other key economists in the immediate post-war period, had anticipated. Rapid European and Japanese economic recovery soon left space for loans entirely to developing countries.

The Bank utilized its multilateral resources in an ingenious way. The capital inputs originally established for member countries – among which the United States initially played a dominant role – required an actual contribution of only 20 per cent of the original capital of the Bank, with the remainder callable. Private financial markets were to be the real source of the money, and this they have remained despite modest increases in capital starting in 1959. By contrast, during its first decades, almost all the IMF's resources came directly from governments, and until the 1970s its focus remained the balance of payments problems of the developed countries.

Private direct investment within the developing countries also received early emphasis through the creation of the International Finance Corporation in 1956. Its performance did not live up to initial hopes. Only since 1995, through direct integration with the Bank, has much greater attention been given to the high cost of typical business procedures in many countries. The publication *Doing Business* has taken on a more significant role as private investment flows have multiplied.

The creation of the International Development Association (IDA) in 1960 fundamentally altered the initial conditions of Bank resources. Its funds, exclusively directed to the poorer countries on favourable terms, required regular triennial contributions. This circumstance allowed periodic legislative discussion of Bank policies, and pressures for policy changes, emanating principally from the United States. Its location in Washington, and the influence of an American president, reinforced that tendency. The initial volume of resources allocated was small. But already by 1980 IDA loans were amounting to equivalent net additions to total Bank lending as the conventional commitments. This part of the story, which was not an issue at the Bank's foundation, has now become the central feature of its decisions. (Other parts of the World Bank, such as the later International Center for the Settlement of Investment Disputes, ICSID, and Multilateral Investment Guarantee Agency, MIGA, should also be noted. These new components corresponded to new functions as the Bank expanded.)

The quality and size of Bank staff, which now totals more than 10,000, also merit mention. Begun at a time when 'development economics' was not yet a part of the economics curriculum within universities, the Bank soon began to employ a talented and professional group that much helped to elevate the status of the subdiscipline. Creation of the *World Development Report* and other publications, regular conference activities, increasingly held abroad, as well as training for many developing country economists through a period at the Bank, have made this intellectual role a positive highlight.

Finally, over a more than 60-year history, the Bank has altered its emphasis dramatically. Now it is best known for its commitment to the elimination of poverty, a subject of little import at its foundation, when its contribution to economic growth was the focus. Now, too, the Bank is exposed to increased opposition from both Left and Right as it searches to retain a role, not only intellectually but also practically. Its net financial contribution has continuously declined as a sophisticated international capital market has expanded, placing more of a burden upon Bank leadership.

In this article, I explore three subjects. First is an assessment of the changing pattern of Bank lending and its effects. Second is an exposition and evaluation of the critiques increasingly directed at the Bank since the early 1990s. Third, by way of conclusion, I raise an essential question: what role should the Bank play in the future?

World Bank Lending: 1946–2005

Table 1 provides the gross and net loan flows, as well as net transfers – with return interest payments subtracted – of the Bank and IDA at decadal intervals since the Bank's foundation. It also presents net private capital flows and net official flows to the developing countries. All have been converted to constant dollars. Three central conclusions immediately follow.

First, the gross real flows of regular disbursements, after an initial acceleration in the 1970s and continuing through to 1985, tend to stabilize thereafter. They then fall off considerably in later years. During the mid-1980s, net flows likewise began to diminish sharply, becoming negative by the early 21st century. Net transfers turn negative shortly after 1985 and become progressively more so thereafter. The Bank has ceased to be a source of resources for middle-income countries some time ago. That is the direct consequence of restricted gross outlays accompanied by amortization and market interest rate charges.

Additionally, the Bank altered its principal mandate in the 1980s, as it had done during the Robert McNamara presidency. Then, in the 1970s, the Bank launched new initiatives to deal with the extensive level of poverty found in the developing world. Income distribution, basic needs, reform of the agricultural sector all figured. With a debt crisis occurring soon after the oil shock of 1979, the Bank underwent a transformation. Structural macroeconomic loans were introduced, moving away from the earlier sectoral emphasis. Conditionality loomed larger, and the Bank began to replicate the simultaneous involvement of the IMF. The Bank became notable for its emphasis upon the primary importance of market signals, as well as privatization and freer trade,

World Bank, Table 1 World Bank lending, total official flows and total private flows, 1960–2005 ($ billion, 2000)[a]

	Fiscal years						
	1960[b]	1970	1980	1990	1995	2000	2005
Disbursements		3.2	11.7	22.4	20.2	18.7	15.8
IBRD	1.8	2.5	8.6	17.0	13.9	13.5	8.7
IDA		0.6	3.1	5.4	6.3	5.2	7.0
Net flows		2.3	9.6	11.5	6.5	6.8	.5
IBRD	1.3	1.6	6.6	6.4	.8	2.9	−4.7
IDA		0.6	3.0	5.1	5.7	3.9	5.2
Net total official flows	22.3	20.0	65.4	69.5	59.2	23.2	−15.2
Net total private flows		20.3	78.4	54.4	226.0	189.7	443.4
Net transfers		1.3	6.0	2.7	−3.0	−2.0	−4.0
IBRD	.9	0.7	3.1	−2.1	−8.1	−5.4	−8.4
IDA		0.6	2.9	4.7	5.2	3.3	4.4

[a]The US chain-type GDP price index (averaging calendar years) has been used. Categories may not sum due to rounding
[b]World Bank loans to developing countries only
Sources: For 1960, Mason and Asher (1973, pp. 208, 219). For other dates, World Bank, *Global Development Finance,* 2001 and 2005

sometimes to the exclusion of a positive role for the state that had figured importantly in the previous decade.

Second, IDA resources, regularly replenished and less diminished by repayments because of their longer term and lesser interest charges, grew substantially. But even these disbursements have not sustained their expansion. Now on a gross basis, they amount to nearly 80 per cent of regular Bank commitments, and will soon exceed them. On a net basis, the IDA proportion is much greater, as can be seen. The disparity is very much stronger when net transfers are recorded. Indeed, it is almost fair to say that the Bank's substantially increased commitment to the alleviation of poverty again in the 1990s was an almost inevitable consequence of its altered resource base. There was no other direction to take.

James Wolfensohn's active presidency, beginning in 1995, was also a major causal factor. Just like McNamara before him, Wolfensohn unleashed new programmes as an advocate of the poor. Coinciding with the rise of India and continuing rapid gains in China, and the beginning of recovery in Africa, his decade of engagement encountered a much better base for this renewed emphasis. Even the initial crises, in Mexico in 1995, in Asia in 1997, and in Russia in 1998 followed by Brazil within months, saw very rapid recovery; this was not a duplication of the lost decade of the 1980s. Wolfensohn transformed the Bank in terms of its managerial style, its relationship with non-governmental organizations (NGOs), its focus on institutions and governance and its emphasis upon concrete results.

Third, the Bank progressively became a marginal contributor of resources over this period, except to the very poor countries. This is clear if one compares the net private flows recorded in Table 1 with Bank, or even total official, lending. To compensate, the Bank's intellectual role has continuously had to be sharpened, redefined and extended, which helps to explain the expansion of branch offices abroad, the relocation of country directors, and the increase in activities within recipient borrowers. This is also why there has been such an emphasis upon information technology as an essential component for spreading knowledge about the development process. What started as a straightforward financial institution has been converted into a far more vocal and innovative participant in the advancement of the position of the poor. The Bank has led in the onerous task of reducing the accumulated debt that burdens many countries. It has also been active in defining a new vision, involving not only governments, but also civil society. But that transformation has not met with universal acclaim.

Should the Bank Survive?

As the Bank approached its 50th anniversary in 1994, increasing unhappiness with its performance became evident. Since the McNamara years, there had been three successive presidents and a changed direction emphasizing macroeconomic programme loans to finance stabilization in a world increasingly adrift. A much broader range of internal reforms, going beyond the balance of payments and domestic savings, was now targeted. Stricter conditions were imposed. Almost 30 per cent of loans were allocated to stabilization objectives, and were concentrated among the highly indebted countries that were suffering from lack of private lending. Unfortunately, the record of accomplishment was not so high. The classification of risk in World Bank documents conveys that reality: in 1970, more than 70 per cent of loans were low risk; in 1980, 30 per cent. Arrears appeared. Positive IDA transfers compensated for negative Bank flows in a number of countries newly eligible for these loans because of their declining income.

Critics multiplied, both on the Left and the Right. For both groups, the status quo was unacceptable. For the more radical, the correct solution was to close the Bank, but for most, the preferred outcome was an altered, more effective institution. Each side envisioned a redesigned Bank more able to accomplish its redesigned objectives.

For the Left, the Bank required nothing less than reinvention – as one critic would later put it – in its operations, concepts and distribution of power. Or, to put the matter another way, the object was a smaller, much more transparent, decentralized and pluralistic development bank. The new institution would be one where developing countries could exercise greater choice and have greater voice. Independent research and significant policy engagement would no longer be features of the Bank. Those functions would devolve to the developing countries themselves.

For the Right, the objective was an equally lesser institution, one that would provide grants to the poorest countries with limited alternative access to financial markets. These funds would be allocated to the conventional objectives identified at the Bank's outset, namely, health care, primary education and infrastructure. No attention would be directed to such issues as the environment, gender equality or labour standards. NGOs, which had become increasingly part of the developing community, would no longer be central participants. Shares of domestic contributions would vary, as a function of per capita income, from ten per cent to 90 per cent. Private capital markets could, and would, substitute for the very modest financial contribution of the Bank to the middle-income countries, and at much lesser cost. The Bank would no longer be engaged in lending to them. Instead, the Bank would limit itself to knowledge transfer and technical assistance for this group of countries. Independent auditors would conduct performance evaluation, emphasizing measured targets and results.

Neither of these directions of reform is now at the centre of discussion. The 60th anniversary has come and gone. One of the important reasons is the impressive acceleration of economic performance within the developing countries since the beginning of the 21st century. Another has been the ability of the Wolfensohn Bank to make itself more acceptable by adopting some of the suggestions from its critics on both sides. Thus the grant element in IDA loans has now risen to 30 per cent (as of 2007); there has been greater attention to the role of the state, as well as the private sector, within developing countries; transparency has increased; and there has been insistence upon country ownership, including broad participation of domestic groups, of the development projects being financed.

This may seem to work in the present. But there remains the question of what lies ahead.

The Future of the Bank

Implicitly, and continuously, the Bank has throughout its history confronted the central issue of whether to give greater weight to growth or equity. During the crisis of the 1980s, the focus temporarily turned to economic recovery. Now, after the triumph of globalization, expansion of

international trade and greater recognition of private sector importance, conditions have seemingly changed. They have also altered in a policy sense, with the establishment of the Millennium goals – UN-mandated objectives in a number of areas that developing countries are supposed to meet by 2015 – and a period of energetic commitment to expansion of social programmes to confront poverty.

As accelerating and generalized expansion has occurred for the first time in three decades, as of the early 21st century more countries are on the verge of graduation, and some of the IDA recipients are approaching their maximum income limit. At the same time, Bank evaluation has raised doubts about some of the new directions pursued: there has seemingly been too much effort directed to the social sectors, governance has not continuously improved nor has corruption been alleviated – despite the importance attached to these issues – and the numbers of people in poverty, other than in Asia, have since the mid-1990s been resistant to improvement. These difficulties are not easy to resolve.

An increase in international private resources as a source for investment has accompanied this expansion. These resources hardly show signs of stopping. Indeed, the speed of global recovery from the tumult at the end of the 1990s is a record accomplishment. Those crises were only a modest pause.

In this new world, the Bank will eventually have to adapt. As the data of Table 1 clearly reveal, neither Bank nor IDA lending is a central source of finance. Although much assistance is granted for political advantage, bilateral overseas development assistance regularly exceeds its net contribution. The Asian and Inter- American Development Banks dominate in their regions.

One direction of change, in a continuation with the recent past, may require an even greater degree of engagement with other agencies of the United Nations, whether the issue is the proliferation of new viruses or the extension of HIV. Another is likely to be a more active participation in global environmental efforts as increasing scientific research indicates the speed and importance of recent climate change. A third may involve efforts to finance infrastructure projects through shared participation of the private and public sectors, with the Bank engaged as a major contributor in poorer countries. A fourth may entail serious accommodation to the implications of changing global supplies of petroleum for the poorest countries.

These are just some possibilities. Many additional ones are sure to emerge. The Bank will have to take on a different role and function. As many developing countries begin to increase their income, the process of graduation implies a change in future leadership at the Bank. No longer will the United States influence choices and policy options as in the past. A new generation of executive directors and employees will debate such future Bank directions internally. External critics will again evaluate whether the Bank should finally cease and desist. Ultimately, however, international institutions have an instinct for survival. The World Bank is probably no exception.

See Also

- ▶ Foreign Aid
- ▶ International Capital Flows
- ▶ International Monetary Fund

Bibliography

Akyüz, Y. 2004. The rationale for multilateral lending: A critical assessment. Unpublished manuscript, July.

Asher, R.B., and M. Edward. 1973. *The World Bank since Bretton Woods*. Washington, DC: Brookings Institution.

Gavin, M., and R. Dani. 1995. The World Bank in historical perspective. *American Economic Review* 85: 329–334.

Gill, I.S., and P. Todd. 2005. *At the frontlines of development: Reflections from the World Bank*. Washington, DC: World Bank.

Kagia, R. 2005. *Balancing the development agenda: The transformation of the World Bank under James D. Wolfensohn, 1995–2005*. Washington, DC: World Bank.

Kapur, D. 2002. *Do as I say not as I do: A critique of G7 proposals on reforming the MDBs*. Working paper No. 16. Washington, DC: Center for Global Development.

Kapur, D., L. John, and W. Richard. 1997. *The World Bank: Its first half-century*. Washington, DC: Brookings Institution.
Krueger, A.O. 1998. Wither the World Bank and the IMF? *Journal of Economic Literature* 36: 1983–2020.
Mason, E.S., and R.B. Asher. 1973. *The World Bank since Bretton Woods*. Washington, DC: Brookings Institution.
Meltzer, A. et al. 2000. *International Financial Institution Advisory Commission report.* Washington, DC.
Mohammed, A.A. 2004. Who pays for the World Bank? G24 discussion paper.
Pincus, J.R., and A.W. Jeffrey. 2002. *Reinventing the World Bank*. Ithaca: Cornell University Press.

World Trade Organization

Robert W. Staiger

Abstract

The success of the General Agreement on Tariffs and Trade (GATT)/World Trade Organization (WTO) as an international institution is widely acknowledged. Among multilateral institutions, the GATT/WTO has adopted a distinctive approach as a forum for international negotiation, based on reciprocal negotiations (over market access) that occur on a voluntary basis between pairs of countries or among small numbers of countries; the results of these bilateral negotiations are then 'multilateralized' to the full GATT/WTO membership under the GATT/WTO principle of non-discrimination. This article describes how recent economic research has attempted to understand and interpret these key design features of the GATT/ WTO.

Keywords

Commitment theories of trade agreements; Cost shifting; Free trade; General Agreement on Tariffs and Trade (GATT); Mercantilism; Most favoured nation (MFN); Non-discrimination in trade; Protection; Reciprocity in trade; Tariffs; Terms-of-trade theories of trade agreements; Unilateral and multilateral trade policies; World Trade Organization (WTO)

JEL Classifications

O5

The World Trade Organization (WTO), like its predecessor the General Agreement on Tariffs and Trade (GATT), has in effect served as the constitution of the post-war international trading system. (The GATT was created in 1947, and the WTO came into existence on 1 January 1995, as a result of the Marrakesh Agreement of April 1994, also known as the WTO Agreement. The WTO Agreement includes the text of GATT: GATT therefore continues to exist as a substantive agreement, but the WTO Agreement also includes a set of additional agreements that build on and extend GATT principles to new areas. Hoekman and Kostecki 1995, provide an excellent institutional overview of GATT and the WTO.) Since 1947, membership in the GATT/WTO has grown from 23 countries to its present size of 150 countries, and according to the WTO's World Trade Report (WTO 2007), average *ad valorem* tariffs on industrial goods have been reduced from upwards of 30 per cent to below four per cent through eight multilateral rounds of negotiation (a ninth, the Doha round, is ongoing at this writing).

The success of the GATT/WTO as an international institution is widely acknowledged. Among multilateral institutions, the GATT/WTO has adopted a distinctive approach to serving as a forum for international negotiation. This approach is based on reciprocal negotiations (over market access) that occur on a voluntary basis between pairs of countries or among small numbers of countries, and the results of these bilateral negotiations are then 'multilateralized' to the full GATT/WTO membership under the GATT/WTO principle of non-discrimination.

As an object of study, the GATT/WTO has attracted the attention of legal scholars since the late 1960s. But until relatively recently, the GATT/WTO has not been the subject of systematic and formal economic analysis. This might seem surprising, because the familiar economic arguments for free trade would seem to provide an obvious foundation for the economic analysis of the GATT/WTO. But this foundation immediately runs into a pair of impediments. First, the

case for free trade is a unilateral case, and it therefore leaves no room for the existence of a trade agreement of any kind: from this starting point, the economic logic of the GATT/WTO is immediately suspect. And second, the liberalizing force that the GATT/WTO has harnessed does not appear to be the consumer gains that come from freer trade: rather, the GATT/WTO is driven by exporter interests. Traditionally, most economists have interpreted these observations as evidence that a mercantilist logic lies at the foundation of the GATT/WTO and that, as a result, economic analysis of the GATT/WTO is futile.

A growing body of theoretical and empirical literature has begun to challenge this view. There are two main branches of this literature (for recent attempts to articulate theories that would constitute a third branch, see Ethier 2006 and Regan 2006). A first branch (terms-of-trade theories) emphasizes the role of trade agreements in providing governments with an avenue of escape from a terms-of-trade driven Prisoner's Dilemma. A second branch (commitment theories) emphasizes the role of trade agreements in providing governments with a means of making commitments to their private sectors. Commitment theories of trade agreements have been developed by a number of authors, and there is also some empirical evidence that the GATT/WTO may play this role (see, for example, Conconi and Perroni 2003; Maggi and Rodriguez-Clare 1998, 2007; and Staiger and Tabellini 1987, 1999). But most of the literature to date adopts the terms-of-trade perspective. So I will focus here on interpreting and evaluating some of the key design features of the GATT/WTO from the perspective of terms-of-trade theories. (Empirical evidence relating to the terms of trade theory of trade agreements is surveyed in Bagwell and Staiger 2002, ch. 11. More recent evidence appears in Broda et al. 2006; and Bagwell and Staiger 2006a.)

All theories of trade agreements must identify a means by which the negotiating governments can gain from the agreement. This entails identifying a 'problem' that would arise in the absence of an agreement, when governments make unilateral trade policy choices. The purpose of a trade agreement can then be viewed as providing a 'solution' to the problem, so that the negotiating governments may share in the associated benefits. The terms-of-trade theory posits that governments can gain from negotiations by correcting the international inefficiencies that occur under unilateral trade policy choices as a result of international cost shifting. This cost shifting arises whenever the government of an importing country increases its import barriers and the prices received by foreign exporters fall as a result, thereby improving the importing country's terms of trade. In this way, a portion of the cost of each government's import protection is borne by foreigners, and as a consequence the unilateral best-response levels of import protection chosen by each government are overly restrictive relative to internationally efficient levels: starting from its best-response (reaction curve) tariffs, each government can therefore gain by negotiating reciprocal liberalization with its trading partners. From the perspective of the terms-of-trade theory, then, the problem associated with unilateral trade policy choices is the cost shifting that importing governments are able to achieve on to foreign exporters; and the purpose of negotiated trade agreements is to give foreign exporters (or their governments) a 'voice' in the trade policy choices of importing governments, so that the 'market access' that each country affords its trading partners can be expanded to internationally efficient levels. (The link between the terms-of-trade theory of trade agreements and the emphasis on market access found in GATT/WTO discussions is identified and formalized in Bagwell and Staiger 2002, ch. 2.)

In this environment, internationally efficient policies can be achieved if each government agrees to adopt the policies it would have chosen had it 'ignored' its ability to shift costs on to foreigners. Accordingly, internationally efficient levels of market access may be delivered under multilateral free trade, but only if all governments seek to maximize national income with their trade policy choices: when governments have broader (for example, political/distributional) goals, international efficiency will generally *not* correspond to free trade. Nevertheless, according to the terms-of-trade theory, the purpose of a trade agreement remains the same independent of government

objectives. This feature suggests that, despite the potential for wide diversity across the objectives of GATT/WTO member governments, the underlying structure of the cost-shifting problem central to the terms-of-trade theory may yield simple and robust insights concerning the logic of key design features of the GATT/WTO.

I now illustrate the basic structure of the international cost-shifting problem at the heart of the terms-of-trade driven Prisoner's Dilemma, and describe how it can account for two pillars of the GATT/WTO: reciprocity and non-discrimination. Broadly speaking, the principle of reciprocity in the GATT/WTO refers to the ideal of mutual changes in trade policy that trigger changes in the volume of each country's imports that are of equal value to changes in the volume of its exports. And according to the non-discrimination principle, a country must provide every other GATT/WTO member country with access to its markets on terms no less favourable than it provides the 'most-favoured' country: hence, under the non-discrimination principle, each GATT/WTO member country faces 'most-favoured-nation' (MFN) tariffs from all other GATT/WTO member countries.

I begin with reciprocity. The essential point can be understood from the perspective of a standard two-country/two-good competitive general equilibrium trade model, in which country A exports good y to country B in exchange for imports of good x. Following Bagwell and Staiger (1999, 2002), government preferences for the two countries can be represented very generally by the functions $W^i(p^i(\tau^i,\tilde{p}^w),\tilde{p}^w)$, where τ^i is 1 plus the ad valorem tariff in country $i \in \{A, B\}$, p^i is the relative price of good x to good y prevailing locally in country i, and $\tilde{p}^w$ is the market-clearing 'world' relative price or terms of trade, which is itself a function of the two tariffs $\tilde{p}^w(\tau^A,\tau^B)$. Under standard conditions $\tilde{p}^w$ is decreasing in τ^A and increasing in τ^B, while p^A is increasing in τ^A and p^B is decreasing in τ^B. Apart from general concavity, the only condition that is imposed on government welfare functions is that $W^A_{\tilde{p}^w} < 0$ and $W^B_{\tilde{p}^w} > 0$, meaning that each government would like more tariff revenue if it could have this without any change in its local prices (and therefore without any change in the distribution or levels of factor incomes within its economy). Because no restrictions are placed on the way in which governments feel about changes in local prices, this representation of government preferences is general enough to include, in addition to the traditional Johnson (1953–1954) national-income maximizing government, the leading models of political economy of trade protection (each of which effectively defines government preferences over redistribution and hence local prices).

The non-cooperative (Nash) tariffs chosen in this environment are defined by the two first-order conditions $W^i_{p^i} + \lambda^i W^i_{\tilde{p}^w} = 0$ for $i \in \{A, B\}$, where $\lambda^i \equiv [\partial \tilde{p}^w / \partial \tau^i]/[dp^i/d\tau^i] < 0$. Notice that international cost shifting is embodied in the term $\lambda^i W^i_{\tilde{p}^w}$ which enters into the first-order conditions, and the presence of this cost-shifting term guarantees that $W^A_{P^A} < 0$ and $W^B_{P^B} > 0$ in the Nash equilibrium. The international efficiency frontier is defined by the (τ^A, τ^B) pairs from which it is not possible to adjust tariffs so as to help one country without hurting the other according to the government preferences W^A and W^B. Formally, this frontier takes the form $\left(1 - \Lambda^A W^A_p\right) = 1/\left(1 - \Lambda^B W^B_p\right)$, where $\Lambda^A\left(1 - \tau^A\lambda^A\right)/\left(W^A_p + \lambda^A W^A_{\tilde{p}^w}\right)$ and $\Lambda^B \left(1 - \lambda^B/\tau^B\right)/\left(W^B_{p^B} + \lambda^B W^B_{\tilde{p}^w}\right)$. From these expressions, a pair of observations can now be confirmed. First, the Nash tariff choices do not achieve the international efficiency frontier, and so there is indeed a 'problem' for an international agreement to solve. And second, politically optimal tariffs, defined by $W^i_{p^i} = 0$ for $i \in \{A, B\}$ and interpreted as the unilateral tariff choices governments would make if they were not motivated by terms-of-trade considerations, do achieve the international efficiency frontier, and so politically optimal tariffs represent a complete 'solution' to this problem. From these observations an important conclusion can be drawn: even in the presence of politically/

distributionally motivated governments, the purpose of a trade agreement is simply to prevent terms-of-trade manipulation.

From this backdrop, we may now ask the question, 'Why would the principle of reciprocity have appealing features?' The answer, simply stated, is that reciprocity describes a fixed-terms-of-trade rule to which mutual tariff changes must conform. (Formally, this can be seen following Bagwell and Staiger 1999, 2002. Define a set of tariff changes $\Delta\tau^A \equiv (\tau_1^A - \tau_0^A)$ and $\Delta\tau^B \equiv (\tau_1^B - \tau_0^B)$ as conforming to reciprocity whenever $\tilde{p}_0^w[M^A(P_1^A, \tilde{p}_1^w) - M^A(P_0^A, \tilde{p}_0^w)] = [E^A(P_1^A, \tilde{p}_1^w) - E^A(P_0^A, \tilde{p}_0^w)]$ where $\tilde{p}_0^w \equiv \tilde{p}^w(\tau_0^A, \tau_0^B), \tilde{p}_1^w \equiv \tilde{p}^w(\tau_1^A, \tau_1^B)$, $p^{0A} \equiv p^A(\tau_0^A, \tau_0^B)$ and $p_1^A \equiv p^A(\tau_1^A, \tau_1^B)$, and where M^A and E^A denote A's imports and exports, respectively. Using balanced trade $(\tilde{p}^w M^A(p^A, \tilde{p}^w) = E^A(p^A, \tilde{p}^w))$, the condition for reciprocity simplifies to the fixed-terms-of-trade rule $[\tilde{p}_1^w - \tilde{p}_0^w] M^A(p_1^A, \tilde{p}_1^w) = 0.)$. And in an environment where terms-of-trade manipulation is *the* problem to be fixed, a fixed-terms-of-trade rule is bound to have some attractive uses. (Bagwell and Staiger 2002, ch. 4 describe and interpret a number of ways in which the principle of reciprocity appears in the GATT/WTO.) Intuitively, the nature of international cost shifting ensures that, beginning from their Nash tariff choices, each government would desire tariff liberalization and the local price movements/greater trade volume that this would bring if this liberalization could be achieved at a fixed terms of trade $\tilde{p}^w$ (that is, recall from above that $W_{P^A}^A < 0$ and $W_{P^B}^B > 0$ at Nash). The principle of reciprocity can be understood to harness this desire, and so to activate efficiency-enhancing tariff-liberalizing forces in this environment.

I now turn to the non-discrimination principle, as embodied in MFN. This requires an extension of the basic two-country model described above to a three-country setting. To this end, let country C have a similar trading pattern to B, in that C also exports good x to country A in exchange for imports of good y. An important feature of the MFN rule is that, in requiring country A to impose a common tariff on imports of x regardless of whether these imports of x originate in exporting country B or C, this rule ensures that a single market-clearing terms of trade $\tilde{p}^w(\tau^A, \tau^B, \tau^C)$ will prevail, and government preferences may continue to be expressed with the simple representation $W^i(p^i(\tau^i, \tilde{p}^w), \tilde{p}^w)$ for $i \in \{A, B, C\}$. Notice that, in the presence of MFN, countries A and B can still negotiate a reciprocal reduction in their respective tariffs τ^A and τ^B that provides each with more trade volume at a fixed terms of trade $\tilde{p}^w$, thereby ensuring that they each gain relative to Nash; and strikingly, as long as A and B abide by reciprocity, there will be no third-party effects of their bilateral negotiation on country C, whose welfare level $W^i(p^C(\tau^C, \tilde{p}^w), \tilde{p}^w)$ remains unaltered owing to the unchanged τ^C and the fixed terms of trade $\tilde{p}^w$. (For C's welfare to remain unchanged, it is in fact not necessary that τ^C remain unchanged, but only that C remain on its tariff reaction curve and that $\tilde{p}^w$ remain unchanged: see Bagwell and Staiger 2006b.) Of course, A and C can engage in bilateral reciprocal negotiations that have the same property. This has an important implication: the MFN rule permits the liberalizing force of reciprocity to be harnessed in an essentially bilateral manner even in a multilateral world. (These and related points are developed in Schwartz and Sykes 1997, and Bagwell and Staiger 2005, 2006b.)

In this general manner, the GATT/WTO pillars of reciprocity and non-discrimination can be understood to underpin the architecture of an international negotiating forum in which the liberalizing force of reciprocity can be harnessed in bilateral negotiations with an assurance of minimal third-party spillovers, thereby permitting each member government – through a sequence of bilateral or small-numbers negotiations – to engineer its escape from a terms-of-trade driven Prisoner's' Dilemma.

See Also

- ▶ Tariffs
- ▶ Trade Policy, Political Economy of

Bibliography

Bagwell, K., and R.W. Staiger. 1999. An economic theory of GATT. *American Economic Review* 89: 215–248.

Bagwell, K., and R.W. Staiger. 2002. *The economics of the world trading system*. Cambridge, MA: MIT Press.

Bagwell, K., and R.W. Staiger. 2005. Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO. *Journal of International Economics* 67: 268–294.

Bagwell, K., and R.W. Staiger. 2006a. What do trade negotiators negotiate about? Empirical evidence from the World Trade Organization. Working paper no. 12727. Cambridge, MA: NBER.

Bagwell, K., and R.W. Staiger. 2006b. *Backward stealing and forward manipulation in the WTO*. Mimeo: Columbia University, Stanford University and University of Wisconsin.

Broda, C., N. Limao, and D. Weinstein. 2006. Optimal tariffs: The evidence. Working paper no. 12033. Cambridge, MA: NBER.

Conconi, P., and C. Perroni. 2003. Self-enforcing international agreements and domestic policy credibility. Working paper no. 114/03. Warwick University Center for the Study of Globalization and Regionalization.

Ethier, W. 2006. The theory of trade policy and trade agreements: A critique. PIER Working paper 06–013. University of Pennsylvania.

Hoekman, B., and M. Kostecki. 1995. *The political economy of the world trading system*. Oxford: Oxford University Press.

Johnson, H.G. 1953–1954. Optimum tariffs and retaliation. *Review of Economic Studies* 1: 142–153.

Maggi, G., and A. Rodriguez-Clare. 1998. The value of trade agreements in the presence of political pressures. *Journal of Political Economy* 106: 574–601.

Maggi, G., and A. Rodriguez-Clare. 2007. A political-economy theory of trade agreements. *American Economic Review* 97: 1374–1406.

Regan, D.H. 2006. What are trade agreements for? Two conflicting stories told by economists, with a lesson for lawyers. *Journal of International Economic Law* 9: 951–988.

Schwartz, W.F., and A.O. Sykes. 1997. The economics of the most-favored-nation clause. In *Economic dimensions in international law: Comparative and empirical perspectives*, ed. J.S. Bhandani and A.O. Sykes. Cambridge: Cambridge University Press.

Staiger, R.W., and G. Tabellini. 1987. Discretionary trade policy and excessive protection. *American Economic Review* 77: 823–837.

Staiger, R.W., and G. Tabellini. 1999. Do GATT rules help governments make domestic commitments? *Economics and Politics* 11: 109–144.

WTO (World Trade Organization). 2007. *World trade report*. Geneva: WTO.

World Wars, Economics of

Stephen Broadberry and Mark Harrison

Abstract

This article focuses on the role of economic factors in explaining the outcomes of the two world wars. In both wars, the scale of resources mobilized was decisive, leaving little room for other factors that feature prominently in narrative accounts, such as national differences in war preparations, war leadership, military organization and morale. The economic advantage of the Allies was not just in size, but also in the quality of their resources, reflected in average real incomes per head of their populations before the wars. We also quantify the economic effects of the wars within a national balance sheet framework.

Keywords

Coalitions; Collectivization; Famine; Fiscal mobilization; Globalization; Lend-Lease; Peasant economy; Rationing; Subsistence; Tariffs; World wars; Economics of the; War and economics

JEL Classifications

H5

The two world wars of the 20th century were events in a single process of reaction against globalization that was prolonged and, from time to time, violent.

From 1815 to 1914 trade and capital flows increased alongside global productivity. Everywhere, economic development tended to reduce local risks. At the same time, falling trade and transport costs exposed farmers, firms, and labouring households to new instabilities and risks that originated far away, in countries and markets across the world. Where governments and politicians embraced these long-range risks,

liberalization fostered engagement in the global economy. Where political entrepreneurs mobilized reaction against them, however, resistance gained ground.

By the end of the 19th century, leaders of several newly industrializing countries were seeking to insulate their economies from global risks through tariff protection. German leaders, for example, aimed at national security through trade within a closed region based on a colonial empire. To secure this empire they launched a naval arms race; the arms race precipitated the formation of two Eurasian alliances that confronted each other in the First World War. On one side stood the Central Powers, primarily the German, Austro-Hungarian and Ottoman empires, joined in 1915 by Bulgaria; on the other side stood the Allies: the British, French and Russian empires, joined in 1915 by Italy, and in 1917 by the United States. But the war brought ruin to the three empires of the Central Powers and to the Russian empire too.

After the First World War, the instabilities intrinsic to the global economic order increased. The weakness of the formerly dominant British economy and the isolation of Germany and Russia undermined global market integration. The slump of 1929 sent deflationary ripples around the world and accelerated the disintegration. As the world market shrank, the great powers struggled over national shares. In the 1930s the world economy broke up into several relatively closed trading blocs. The British, French and Dutch reorganized their trade on colonial lines. With Hitler in power, Germany resumed the perspective of regulated trade within a colonial region in central and eastern Europe, and this led to rivalry with other interested regional powers. Italy established bilateral trade with the smaller states of the former Austro-Hungarian empire, and also set about winning an African empire. The Japanese competed with the Americans, the Dutch, the British and the Soviet Union for influence in east Asia and the Pacific. The Soviet Union developed a closed economic space behind the frontiers of the former Russian empire, and defended it against the Japanese.

The worldwide trade disintegration contributed to the causes of the Second World War. The economies of the Axis powers, Germany, Italy and Japan, were too small to prosper without specialization and external sources of food, fuel and other materials. A common thread in their course of external aggression was the attempt to secure these supplies by imposing a colonial regime upon trading partners. In this sense the national policies that led to the Second World War were a continuation of those that had led to the First World War.

The second war continued some of the themes of the first, but it was not just a repeat. The object of the first war was regional – to control Europe, the Atlantic and the Near East. The second war was a struggle for global domination in the full sense. The first war was certainly fuelled by racial identities, but no one aimed at genocide, as they did in the second. The first war ended inconclusively, with a ceasefire and a peace treaty that tried to punish the aggressors, but there was no unconditional surrender, and in Germany those who wanted to try again eventually took power. In 1945 the aggressors were crushed militarily and morally, their surviving leaders were put on trial, and what they stood for was excluded from public life.

In this article we pursue the similarities and differences of the two wars in terms of economic history. We have two main themes. First, what is the power of the economic factors compared with others in helping to explain the outcomes of the two wars? Second, of the possible economic factors that should be considered, which contribute most to the explanation of the results? These are not new questions, of course; here, we outline briefly some alternative views.

Historians of the two world wars tend to narrate their story as a complex interplay of forces that worked at many levels. They tell a story of warfare that was increasingly mechanized and waged for years on end by massed forces. Nonetheless, war was waged by people, not by numbers. Economists, in contrast, have tended to give the centre stage to the numbers, conceding less to aspects of warfare such as leadership, discipline,

heroism and villainy. Raymond Goldsmith (1946, p. 69), an economist who helped to manage the United States war economy in his youth, once observed that:

> The cold figures of the output of airplanes, tanks, guns, naval ships, and ammunition, particularly when they are reduced to the still colder form of indices of aggregate munitions production of the major belligerents, probably tell the story of [the Second World War] as well as extended discussion or elaborate pictures... They back to the full the thesis, dear to the economist's ear, that whatever may have saved the United Nations from defeat in the earlier phases of the conflict, what won the war for them in the end was their ability – and particularly that of the United States – to produce more, and vastly more, munitions than the Axis.

To many historians this view remains unappealing; Richard Overy (1998), for example, has objected that it leaves no room for 'a whole series of contingent factors– moral, political, technical, and organizational – [that] worked to a greater or lesser degree on national war efforts'.

The opposition between cold figures and hot blood is false to some extent. Of course, leadership and psychology mattered. But they mattered less than in previous eras because they had become problems that both sides could solve. In both world wars, multi-million armies took the field and stayed there for months and years, giving and taking appalling losses, without disintegrating. Since the moral fabric of military life could withstand the pressure, numbers of men and the volume of supplies assumed the decisive role.

If economics did matter, exactly what was it about the economies of the Allies that gave them superiority? In Goldsmith's tradition, size mattered and only size. Niall Ferguson is a historian who gives economics the attention it deserves. Noting the overwhelming size advantage of the Allies in population and production on the eve of the First World War, he remarks (1998, p. 248), 'To the economic historian, the outcome of the First World War looks to have been inevitable from the moment [the British] opted for intervention'. Given this advantage, he argues, the war should have been over quickly; the only explanation for the Allied failure to conclude the war much sooner is Allied mismanagement, so Ferguson concludes that the Allied economic preponderance was 'an advantage squandered'. As a result, economic advantage came into play only after much time had passed and the military advantage of the aggressors had almost won the day.

There is much truth in this, but we will take a more nuanced view of what it was about economic life that could be decisive in warfare. The belligerents' economies differed not only in the volume of national resources but also in their quality. The main factor in quality was the level of peacetime economic development, which we measure by average real incomes per head of the population. Richer countries could mobilize production, public finance, soldiers and weapons out of proportion to their general economic capacities; in other words, the level of economic development acted as a multiplier of size. For Britain in both world wars, control of the vast but impoverished population and territory of India, for example, mattered little compared with access to the rich markets of the United States.

The First World War

From an economic viewpoint, the First World War can be divided into two phases. In the late summer of 1914, both sides hoped for a quick victory with a limited commitment of resources to the war effort. This first phase is summed up in the memorable phrase 'business as usual', which was common currency in Britain at the time (Lloyd 1924). It was hoped that the war could be fought along similar lines to previous centuries, with a clear distinction between soldiers doing the fighting and civilians getting on with normal life. However, from late 1916 both Britain and Germany stepped up mobilization in the direction of 'total war'. In total war, industry was mobilized to provide unprecedented amounts of munitions, and industrial workers became as vital to the war effort as soldiers. During this second phase, keeping up production and avoiding economic collapse became central to management of the war. The first economy to collapse was on the Allied side;

the Bolshevik Revolution of 1917 took Russia out of the war and led to a Soviet republic (Gatrell 2005). In 1918, falling output in Turkey, Austria and Germany led to the collapse of the Central Powers and the break-up of their empires (Pamuk 2005; Schulze 2005; Ritschl 2005). France also suffered a late collapse of output, but was shored up by the other Allies (Hautcoeur 2005).

Size and Development

The Allies mobilized more soldiers and produced more of most weapons than the Central Powers, as can be seen in Table 1. Furthermore, the degree of Allied superiority increased with the complexity of the weapons. Only in guns did the Central Powers have numerical superiority, while the Allied superiority in tanks reached a factor of nearly 90:1.

Table 2 shows that the balance between the two sides varied over time, as the alliances' compositions changed. In 1914, the Triple Entente of the United Kingdom, France and Russia could also draw on their colonies, and were joined by other countries including Serbia, the British Dominions, Liberia and Japan. By November 1916, the Allies had been joined by a second wave of countries, including Italy, Portugal and Rumania. By November 1918, although Russia had dropped out, the Allies had been strengthened by the United States and a further wave of countries. By this time, the Allied side included 70 per cent of the world's pre-war population and 64 per cent of its pre-war output. The scale of resources that could be mobilized by the Central Powers varied less over time. Austria-Hungary started the war, joined immediately by Germany and shortly after by the Ottoman Empire. By November 1915, Bulgaria had also joined, but Italy, defaulting on its treaty obligations, joined the Allies.

It is important to consider the level of economic development of individual countries as well as the volume of output that the two alliances could draw upon. Britain, for example, with a prewar population of 46 million, had average incomes of nearly $5,000 (at 1990 prices), but its colonies, excluding the Dominions, had a pre-war population of 380 million, mostly in India, with average incomes of less than $700. Thus the colonies, with nearly eight times the population of Britain, produced only about the same volume of output. But the colonial output was less available for fighting the Germans because most of it was needed to meet the subsistence needs of the colonial population. Furthermore, this population was difficult to mobilize because of its distance from the theatre of war and the level of development of colonial administration. Even within the Triple Entente, the low level of development in Russia limited the Allied mobilization. The Central Powers were similarly hampered by the low level of development of the Ottoman Empire and Bulgaria, and even the Hungarian half of the Habsburg Empire.

World Wars, Economics of, Table 1 Allies vs. Central Powers: soldiers and equipment in the First World War

	Allies (1)	Central powers (2)	Ratio, 1:2 (3)
Soldiers mobilized, million	41.0	25.6	1.6
Weapons produced:			
Guns, thousand	59.9	82.4	0.7
Rifles, million	13.3	12.1	1.1
Machine guns, thousand	656	319	2.1
Aircraft, thousand	124.5	47.3	2.6
Tanks	8,919	100	89.2

Source: Broadberry and Harrison (2005)

By comparing the information for the two alliances in Table 2, it is possible to calculate size and development ratios for three benchmark dates in Table 3. The ratios are calculated for each alliance as a whole, and also for the great powers only. The rationale for the latter is that if, as we argue, poor colonies did not count for much, it is helpful to see how the ratios look if we do not count them at all. The table establishes a striking result: judging by economic size, the Central Powers were doomed to defeat. In November 1914, the Allies had access to five times the population, 11 times the territory and three times the output of the Central Powers. If we look only at great powers, the Allied advantages in population and output were

World Wars, Economics of, Table 2 The alliances in the First World War: resources of 1913

	Population, million	Territory		GDP in 1990 prices	
		Million sq. km	Ha. per head	$ billion	Per head, $
Allies					
November 1914					
Allies, total	793.3	67.5	8.5	1,093.6	1379
UK, France and Russia only	259.0	22.6	8.7	622.8	2405
November 1916					
Allies, total	853.3	72.5	8.5	1,210.5	1419
UK, France and Russia only	259.0	22.6	8.7	622.8	2405
November 1918					
Allies, total	1,271.7	80.9	6.4	1,760.6	1384
UK, France and USA only	182.3	8.7	4.8	876.6	4809
Central powers					
November 1914					
Central powers, total	151.3	5.9	3.9	376.6	2489
Germany and Austria-Hungary only	117.6	1.2	1.0	344.8	2933
November 1915					
Central powers, total	156.1	6.0	3.8	383.9	2459

Source: Broadberry and Harrison (2005)

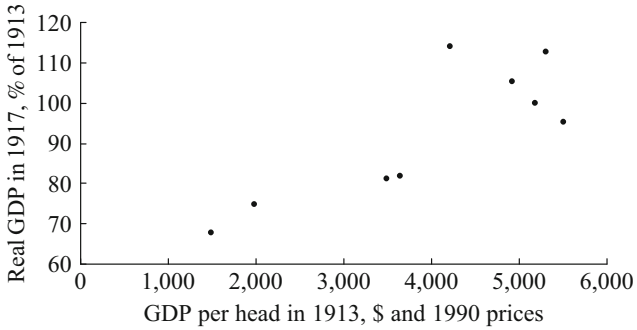
World Wars, Economics of, Table 3 Allies versus central powers: size and development ratios

	Population	Territory	Territory per head	Gross domestic product	GDP per head
November 1914					
Total	5.2	11.5	2.2	2.9	0.6
Great powers only	2.2	19.4	8.8	1.8	0.8
November 1916					
Total	5.5	12.1	2.2	3.2	0.6
Great powers only	2.2	19.4	8.8	1.8	0.8
November 1918					
Total	8.2	13.5	1.7	4.6	0.6
Great powers only	1.6	7.5	4.8	2.5	1.6

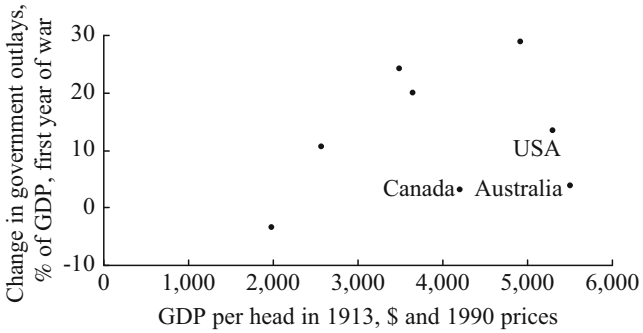
Source: Calculated from Table 1

smaller, but larger in territory, reflecting the fact that German and Turkish colonies tended to be in the sandy deserts of Africa and the Middle East. However, the Allied advantage was limited by relatively low average incomes in Russia and the British and French colonies. Allied incomes were less than two-thirds the average level of the Central Powers, or 80 per cent if attention is confined to the great powers, if Russia is counted as a great power.

By November 1916 the Allied advantage had grown moderately in terms of population, territory and output, but the Central Powers continued to have an advantage in average incomes. By November 1918, however, the situation had changed dramatically, largely as a result of the United States replacing Russia. The Allied advantages in population, territory and output all increased markedly, and for the first time the Allies enjoyed an average income advantage if attention is restricted to great powers. Although it took some time for the American presence to be felt on the battlefield, it sealed the Central Powers' fate.



World Wars, Economics of, Fig. 1 Production mobilization: nine countries, 1913–1917 (*Notes*: observations are, from left to right, Russia, Austria-Hungary, France, Germany, Canada, UK, New Zealand, USA and Australia. *Source*: Broadberry and Harrison (2005))

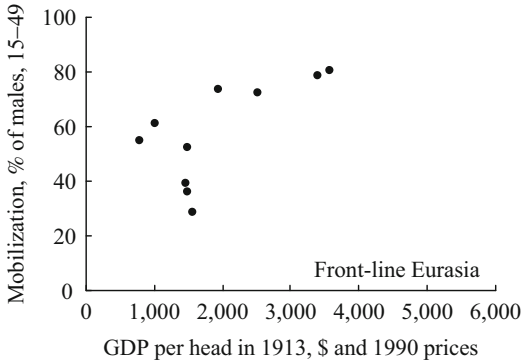


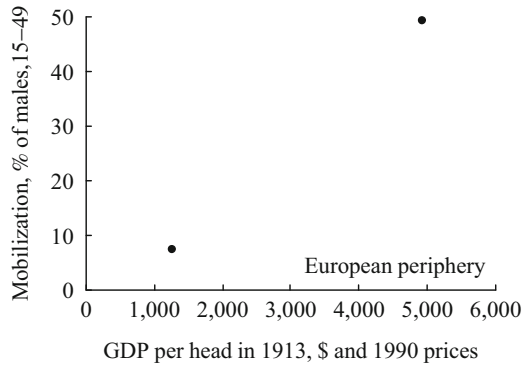
World Wars, Economics of, Fig. 2 Fiscal mobilization in the First World War: eight countries (*Notes*: observations not labelled within the figure are, from left to right, Austria-Hungary, Italy, France, Germany and UK. *Source*: Broadberry and Harrison (2005), supplemented by Austria-Hungary from Schulze (2005))

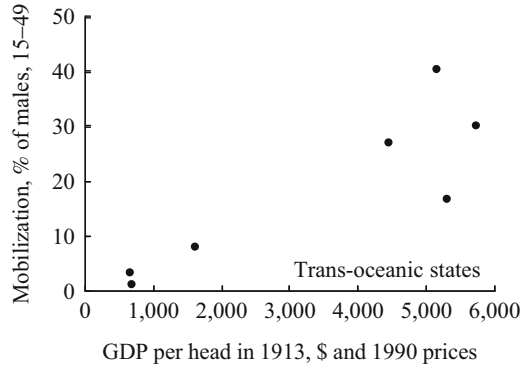
Development and Mobilization

The ratios in Table 3 are based on the assumption that during the war the real output of a given country did not change. The reason for this assumption is statistical: it is difficult to track GDP changes in wartime in the poorer countries. What information we have suggests that Table 3 must understate the actual swing in favour of the Allies during the war, because output increased in the United States and Britain but fell in the less developed economies of the Central Powers. This can be seen in Fig. 1, which plots the change in GDP during 1913–17 against the level of per capita income in 1913 for nine countries. The relationship is strongly positive, reflecting the fact that rich countries were better able to mobilize output than poor countries. The biggest decline was in Russia, which was also the poorest amongst these countries in 1913, and collapsed in the Bolshevik Revolution of 1917. There were also substantial falls in output in Austria-Hungary, France and Germany, which all started the war with lower average incomes than Canada, Britain, New Zealand and the United States.

Another measure of mobilization which varied with the level of development is the change in government outlays as a share of GDP. This reflects the extent to which governments were able to convert output from peacetime uses to the war effort through taxation and spending. Figure 2 plots this measure of fiscal mobilization during the first year of the war against pre-war average incomes for eight countries. The relationship is positive, but not as clear cut as for production mobilization. In particular, it is necessary to control for distance from the main theatre of war in Europe, with the New World countries of Canada, Australia and the United States mobilizing fewer resources through taxation and public spending than the European countries.







World Wars, Economics of, Fig. 3 Military mobilization in the First World War: 18 countries and the French colonies (*Note*: observations are, from left to right: Front line Eurasia: Serbia, Turkey, Russia, Bulgaria, Roumania, Greece, Austria-Hungary, Italy, France, and Germany. European periphery: Portugal and UK. Non-European states: French colonies, India, South Africa, Canada, New Zealand, USA, Australia. *Sources*: GDPs per head in 1913 from Tables 1 and 2 or, if not listed there, from Maddison (2001, p. 185); cumulative mobilization rates, 1914–1918, from Urlanis (1971, p. 209))

Perhaps the most direct measure of mobilization is the share of the prime-age male population recruited into the military. This measure, plotted in Fig. 3 against average incomes in 1913, is available for a relatively large sample of countries. Again, we find a relationship that increases with pre-war prosperity and decreases with distance from the main theatre of war. The figure is plotted in three distance bands, comprising the frontline Eurasian states, peripheral European countries isolated from the frontline by land or sea (Britain and Portugal), and non-European states. Cumulative numbers mobilized are shown as a proportion of males aged 15–49. After we have controlled for distance (that is, within each distance band), there is a positive relationship between military mobilization and the level of development. But dropping a band also lowers the mobilization rate substantially.

Figures 1, 2, and 3 show us that the level of development acted as a multiplier of size. Rich countries were able to mobilize production, public finance and soldiers out of proportion to the size of their economies measured by GDP.

Mobilization and Agriculture

Why did being poor matter for large countries like Russia, Austria-Hungary and Turkey? During the First World War, the answer can be found in the performance of the agricultural sector, since these countries all ran short of food long before they ran out of guns and shells (Offer 1989). Broadberry and Harrison (2005) attribute this to the negative impact of peasant agriculture on mobilization.

One of the most striking attributes of relative poverty was the role of subsistence farming.

Contemporary observers were aware of these differences and interpreted them as follows: when war broke out, a country such as Russia would have an immediate advantage in that most of the people could feed themselves; moreover, the diversion of food supplies from export to the home market would actually increase Russia's advantage. In contrast, Britain would quickly starve (Gatrell and Harrison 1993). This diagnosis could not have been more wrong. In practice a large peasantry proved to be a great disadvantage in mobilizing resources for war. Meyendorff (cited by Gatrell 2005) described what happened in Russia as 'the Russian peasant's secession from the economic fabric of the nation'. And not only from Russia, for Italy, Austria-Hungary, the Ottoman Empire and Germany all had large peasant populations that proved extremely difficult to mobilize for much the same reason. In wartime peasant agriculture behaved like a neutral trading partner. Why should the Netherlands trade with Germany at war, given the latter's reduced ability to pay, except under threat of invasion and confiscation? Peasant farmers, trading with their own governments, made the same calculation. Thus the Russian economy looked large, but if the observers of the time had first subtracted its peasant population and farming resources they would have seen how small and weak Russia really was.

The peasant's propensity to secede is clearly visible from a comparison of the richer and poorer countries' experience. When war broke out British and American farmers were offered higher prices, responded normally to incentives, and boosted production. The fact that British farming had already contracted to a small part of the economy made its expansion easier: there were plentiful reserves of land unused or little exploited, and the high productivity of farm labour meant that large increases in farm output could be achieved with few additional resources (Olson 1963; Broadberry and Howlett 2005).

In the poorer countries, in contrast, wartime mobilization took resources away from farming, particularly young men and horses for the army. Once in the army these young men and horses still needed to be fed, which required a diversion of food supplies from rural households to government purchasers. But the motivation for farmers in the countryside to sell food was reduced, not increased. These were subsistence farmers who grew food partly for their own consumption; what they sold, they took to the market primarily to buy manufactured commodities for their families. But war dried up the supply of manufactures to the countryside. The small industrial sectors of the poorer countries were soon wholly concentrated on supplying the army with weapons and kit. Little capacity was left to supply the countryside, which faced a steep decline in supplies.

Consequently, peasant farmers retreated into subsistence activities and the economy began to disintegrate. There might still be plenty of food, but it was locked in the countryside. The farmers preferred to eat it themselves than sell it for a low return. What food it could get, the government gave to the army for a simple reason: hungry soldiers will not fight. Between the army and the peasantry the urban workers were caught in a double squeeze. As the market supply of food dried up, urban food prices soared, and an urban famine set in. In terms of the economics of famines, the primary cause was not a failure of production but the urban society's loss of food entitlements (Sen 1983; Offer 1989).

Aware of this, public opinion might blame unpatriotic speculators or incompetent officials. But the truth was that a poor country had few genuine choices. The scope for policy to improve the situation was more apparent than real, and government action often made things worse: the Russian, Austrian and German governments all began to ration food to the urban population, for example, while attempting to buy food from the farmers at purchasing prices that were fixed low for budgetary reasons. To repeat: in richer countries the government paid *more* to the farmers, and this worked, but in poorer countries the government tried to pay *less* and this had entirely predictable results: the farmers' willingness to participate in the market was further undermined.

In summary, in wartime poor countries suffered the consequences of peasant agriculture, which was essentially a deadweight on their mobilization efforts. Economic mobilization led to

World Wars, Economics of, Table 4 Destruction of human and physical capital (per cent of pre-war assets)

	Human capital	Physical capital			
		Domestic assets	Overseas assets	Reparations bill	National wealth
Allies					
Britain	3.6	9.9	23.9	...	14.9
France	7.2	24.6	49.0	...	31.0
Russia	2.3	14.3	...	...	...
Italy	3.8	15.9	...	...	...
United States	0.3	...	...	...	...
Central Powers					
Germany	6.3	3.1	...	51.6	54.7
Austria-Hungary	4.5	6.5	...	...	...
Turkey and Bulgaria	6.8	...	...	...	...

Source: Derived from Broadberry and Harrison (2005, p. 28)

urban famine, revolutionary insurrection, and the downfall of emperors in Russia, Austria-Hungary, Germany and Turkey. The same process began in France, which still had a large peasant sector in 1914, but Allied support nipped it in the bud.

War Losses

After the First World War, there were several attempts to calculate the costs of the war. However, these studies fell out of fashion, tainted by association with inflated demands for reparations, and because later writers became interested in any positive developments that could be identified as arising out of the carnage and destruction. Thus in his popular survey of the First World War, Hardach (1987, p. 286) argues that Bogart's (1920) estimates of the costs of the war have not been revised in the light of later evidence because 'the whole basis of calculation has been recognized as inappropriate'.

There are good grounds to be sceptical, however, about the revisionist view that associates war with accelerated economic development. Milward (1984, pp. 17–18), a leading revisionist, cites Bowley (1930) as a pioneer of revisionism, but Bowley himself (1930, pp. 21–3) pointed out how difficult it is to show that any wider changes were actually the result of the war and would not have occurred anyway in its absence. Classifying developments as (*a*) mainly unconnected with the war, (*b*) accelerated or retarded by it or (*c*) apparently arising out of it, Bowley was himself reluctant to put anything other than the key elements of Bogart's 'cost of war' calculations such as loss of life and destruction of capital into category (*c*). He did mention the new economic relations between Europe and the United States in this category, but with hindsight we can see that the process of US overtaking was already under way well before 1914 (Abramovitz 1986; Broadberry 1998).

Table 4, accordingly, provides updated estimates of the destruction of human and physical capital as the costs of war within a national balance sheet framework provided by Broadberry and Howlett (1998). The first element, the destruction of human capital, is measured by war deaths relative to the population aged 15–49. This differs from the true proportion of human capital destroyed by the war to the extent that younger cohorts had more human capital investment, particularly through education. Although Germany suffered the most casualties in absolute numbers, the proportionate losses were higher in France, Serbia-Montenegro and Roumania among the Allies, and in Turkey and Bulgaria among the Central Powers.

The domestic physical capital losses in Table 4 build upon the work of Bogart (1920), who estimated property losses on land and shipping and cargo losses. These are expressed as a proportion of physical capital from modern historical national accounting sources. The French figures draw on estimates of losses from the Reparations Commission and capital stock data from Carré et al. (1976, p. 151). Although these probably overstate French losses, alternative estimates by

Villa (1993) yield implausibly low ratios, given the concentration of fighting on French soil (Hautcoeur 2005, p. 199). Russia's losses were proportionately high, more because of the small size of the pre-war capital stock than a large absolute amount of wartime destruction.

For some countries in Table 4, we can estimate the change in overseas assets and national wealth. Nearly a quarter of British overseas investments was liquidated during the war, so that the reduction of national wealth was proportionally much greater than the loss of physical capital. The loss of French overseas assets was proportionally very high due to heavy exposure to Russian loans, so that, as in Britain, the share of national wealth lost in the war was proportionally greater than the share of physical capital lost.

Finally in Table 4, we have added in Germany's reparations bill as a proportion of pre-war capital, since it represented an increase in foreign liabilities and hence a reduction in national wealth just as much as the liquidation of Britain's overseas assets meant a reduction in national wealth. Of course the extent to which Germany actually had to pay these reparations is much debated, but that does not alter the effect on the national balance sheet as it stood in 1919, immediately after the Treaty of Versailles. These figures include the A + B + C Bonds, which added up to a total of 132 billion Gold Marks.

The Second World War

Like its forebear, the Second World War may be divided into two periods. In the first period, economic considerations were less important than purely military factors. This was the phase of greater success for the powers of the Axis, and it lasted from 1937 when the war began in the Pacific, or from 1939 in Europe, until the end of 1941 or 1942; the exact turning point differed by a few months among the different regional theatres. In this first period, Germany and Japan had advantages of strategy and fighting power on their side. As a result, they were able to inflict overwhelming defeats upon an economically superior combination of powers. In early 1942, Richard Overy writes (1995, p. 15), 'no rational man... would have guessed at the eventual outcome of the war'.

This phase ended, however, without the decisive victory that previously appeared within the Axis powers' grasp. What ended it? On the surface it was the military failures, not economic weaknesses, of the Axis. Beneath the surface, however, economic fundamentals reasserted themselves: while the Allies had given ground everywhere, their morale had stiffened, their economies were not exhausted, their cooperation was taking effect, and their industries were supplying the front with a rising flood of munitions that would eventually overwhelm the adversary.

In the second period of the war, which began in 1942, the early advantages of the Axis evaporated. There was a brief stalemate. A war of attrition developed in which the opposing forces ground each other down, with rising force levels and losses. Superior military qualities came to count for less than superior size, wealth and economic mobilization. Economic superiority let the Allies take risks, absorb the cost of mistakes, replace losses, and accumulate overwhelming force. This turned the balance against the Axis and won the war.

This narrative does not support the claim that only economics mattered. Economic factors were decisive, however, in the context of a simple fact. The Axis leaders had the chance to use the other factors to decide the war, and they failed. Their failure gave the Allies the chance to bring economics decisively into the equation.

Size and Development

Table 5 shows the volumes of combat resources that each side delivered to the theatres of the Second World War. A comparison of the totals with those in Table 1 shows a staggering increase: a quarter of a million tanks and half a million aircraft, for example, compared with 170,000 aircraft and fewer than 10,000 tanks in the First World War. One thing remained the same, however, across the two wars: the Allies supplied a greater volume of combat resources than their combined adversaries in almost every respect.

The Allied advantage did not hold at all points of time and place. As Goldsmith remarked, the pre-war rearmament of the Axis powers gave

World Wars, Economics of, Table 5 Allies vs Axis: soldiers and equipment in the Second World War

	Allies (1)	Axis (2)	Ratio, 1:2 (3)
Combatant-years, million	106.4	76.9	1.4
Weapons Produced:			
Rifles and carbines, million	25.3	13.0	1.9
Combat aircraft, thousand	370	144	2.6
Machine guns, thousand	4,827	1,646	2.9
Guns, thousand	1357	462	2.9
Armoured vehicles, thousand	216	51	4.3
Mortars, thousand	516	100	5.1
Major naval vessels	8,999	1,734	5.2
Machine pistols, thousand	11,604	1,185	9.8
Ballistic missiles	0	6,000	...
Atomic weapons	4	0	...

Source: Harrison (1998, pp. 14–16) except that numbers in the French armed forces in 1940 are corrected as noted by Harrison (2005). The number of ballistic missiles is an approximate upper limit based on Ordway and Sharpe (1979, pp. 405–7). Of the four bombs produced by the Manhattan Project one was tested at Alamogordo, two were exploded over Japanese cities, and one remained unused

them an early start and this, combined with their purely military advantages, accounts for their early success. A balance struck at the end of 1940, for example, when France had dropped out, the United States remained neutral, and the Soviet Union was still Germany's partner of convenience, would show a picture of Allied disadvantage. By 1942, however, reinforced by America and Russia, the Allies outnumbered and outgunned the powers that they faced in every major theatre. This was true even on the eastern front where Germany and the USSR confronted each other. These two powers were of similar economic size measured by GDP and industrial production, but the Soviet Union was substantially poorer in terms of the average incomes of its much larger population. Although this disadvantage was enlarged by devastating military and territorial losses in 1941 and 1942, the Soviet Union fielded a bigger army and supplied it more generously. We return to this anomaly below.

The relative economic sizes of the powers and their colonial possessions are shown in Tables 6 and 7. If we consider the world as it was on the eve of the war, then the populations available to the Allies – principally Britain and France with their colonies and dominions, but also including Poland and Czechoslovakia – amounted to nearly 690 million people occupying nearly 48 million square kilometres. The total output of this territory is estimated at one trillion dollars in 1990 prices. Against them stood the nearly 260 million people and more than six million square kilometres available to the Axis powers – Germany, Italy and Japan, with the Italian and Japanese empires and annexations at the time such as Austria, Korea and Manchuria. The people and their lands on the side of the Allies exceeded those available to the Axis power by several times.

The size of this advantage is more statistical than real, although a real advantage remains after the statistics are stripped out. Africa and South Asia, poor, undeveloped and relatively sparsely settled, made up the greater part of the Allied advantage in size. When we turn to total output, it turns out that the Allied GDP exceeded that of the Axis territories by only one-third; this is because average incomes across the Allied territories – less than $1,500 in modern prices – stood at only one half the $2,900 level of the Axis territories. Here is an ironic comment on the colonial aspirations of the Axis powers: what they wanted so much, and did not yet have, was access to millions of square kilometres of poorly integrated, low-yielding farmland and remote semi-desert.

As before, since poor colonies did not count for very much, we also count the resources on either side considering the great powers only. The Allied size advantage now disappears since Germany, Italy and Japan together had twice the population and one and a half times the territory of Britain and France – but it is replaced by a development advantage: the GDP per head of the Allied powers exceeded that of the Axis powers by one half.

Tables 6 and 7 also show how this balance evolved from 1938 to 1942, when the domains under control of the Axis powers had reached their maximal extent. As their forces swept across Europe and the Pacific region the population

World Wars, Economics of, Table 6 The alliances in the Second World War: resources of 1938

	Population, million	Territory		GDP in 1990 prices	
		Million sq. km	Ha. per head	$ billion	Per head, $
Allies					
1938					
Allies, total	689.7	47.6	6.9	1,024	1,485
UK and France only	89.5	0.8	0.9	470	5,252
1942					
Allies, total	783.5	68.0	8.7	1,749	2,232
UK, USA and USSR only	345.0	29.3	8.5	1,444	4,184
Axis					
1938					
Axis, total	258.9	6.3	2.4	751	2,902
Germany, Austria, Italy and Japan only	190.6	1.2	0.7	686	3,598
1942					
Axis, total	634.6	11.2	1.8	1,552	2,446
Germany, Austria, Italy and Japan only	190.6	1.2	0.7	686	3,598

Source: Harrison (1998, pp. 3–9)

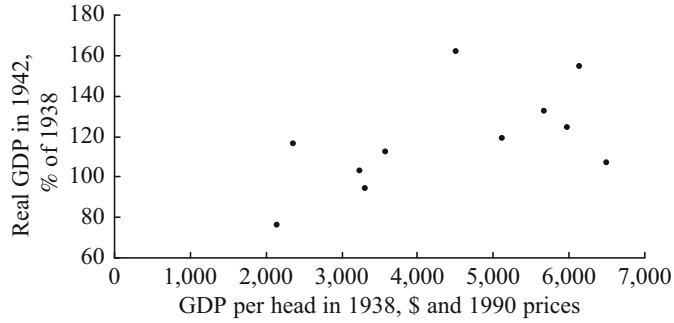
World Wars, Economics of, Table 7 Allies versus Axis: size and development ratios

	Population	Territory	Territory per head	Gross domestic product	GDP per head
1938					
Total	2.7	7.5	2.8	1.4	0.5
Great Powers only	0.5	0.6	1.4	0.7	1.5
1942					
Total	1.2	6.1	4.9	1.1	0.9
Great Powers only	1.8	23.5	13.0	2.1	1.2

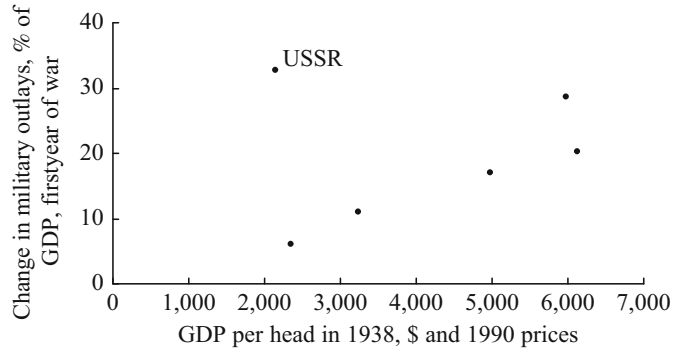
Source: Calculated from Table 6

under Axis control tripled, while territory and peacetime output potential doubled; the addition of hundreds of millions of east European and east Asian farmers led the average development of the Axis empires to decline somewhat, from $2,900 to less than $2,500 in modern prices. Britain, in contrast, lost its allies France – with its empire –and Poland, but was joined by the Soviet Union and the United States. Between 1938 and 1942, therefore, the Allied population and territorial sway increased somewhat, but its peacetime output rose by three quarters (from one trillion to 1.75 trillion dollars). This is *before* the wartime doubling of United States output is taken into account. Joining the richest and poorest of the great powers into a single coalition had a mixed effect, of course, but the net result was an increase in the measured average level of development across all the Allied territories from less than $1,500 in modern prices to more than $2,200.

Table 7 converts these figures into ratios of Allied advantage or disadvantage. We see that in 1942, when things were at their worst, the Allied powers alone had nearly twice the population, more than twice the output, and more than 20 times the territory of the Axis powers. All they had to do was not to lose; given enough time, this economic preponderance would surely bring victory. The weakest link in the Allied chain was poor Russia, with its hundred million low-productivity peasants and seven million square kilometres of permafrost. Germany had forced Russia out of the First World War; could the same not happen again?



World Wars, Economics of, Fig. 4 Production mobilization: 11 countries, 1938–1942 (*Note*: observations are, from left to right, the Soviet Union, Japan, Italy, Finland, Austria, Canada, Germany (excluding Austria), Australia, UK, USA and New Zealand. *Sources*: Harrison (1998, 10) and Maddison (2001); for Soviet GDP see also the sources listed under Table 5)



World Wars, Economics of, Fig. 5 Fiscal mobilization in the Second World War: six countries (*Notes*: observations are, from left to right, the Soviet Union, Japan, Italy, Germany, the UK and the USA. *Source*: Harrison (1998, p. 21))

Development and Mobilization

As with the First World War we will consider three dimensions of mobilization: production (the increase in total output that was achieved during the war), the government's fiscal leverage (the extent that output was mobilized through government spending and taxation out of peacetime uses into the war effort), and military mobilization (the degree of mobilization of the population into uniform). Each of these was powerfully influenced by the pre-war level of development of the economy.

Figure 4 shows production mobilization plotted against pre-war average income. Under the pressure of war, rich countries expanded their economies; poor countries tended to collapse, and the collapse proceeded further, the poorer they were. Figure 5 shows the speed of fiscal mobilization. The slope of the relationship with pre-war economic development has the same positive sign: only rich countries achieved significant fiscal mobilization, but there is an exception: the Soviet Union. Some underlying figures are provided in Table 8: these confirm that the Soviet Union achieved a level of mobilization of GDP into the war effort – three-fifths at its peak– that was typical of much richer countries. Germany and Japan achieved similar degrees of mobilization only in the last spasm of the struggle that preceded immediate collapse and defeat.

The Soviet anomaly demands explanation. A relatively poor country, Russia collapsed in the First World War, and the Soviet Union could have been expected to do so again in the Second

World Wars, Economics of, Table 8 The military burden 1939–44 (military outlays, per cent of national income)

	1939	1940	1941	1942	1943	1944
At current prices						
Allied powers						
USA	1	2	11	31	42	42
UK	15	44	53	52	55	53
USSR	...	...	...	...	...	...
Axis powers						
Germany	23	40	52	64	70	...
Italy	8	12	23	22	21	...
Japan	22	22	27	33	43	76
At constant prices						
Allied powers						
USA	1	2	11	32	43	45
UK	...	...	...	...	...	...
USSR	...	17	28	61	61	53
Axis powers						
Germany	23	40	52	63	70	...
Italy	...	...	...	...	...	...
Japan	...	...	...	...	...	...

Source: Harrison (1998, p. 21)

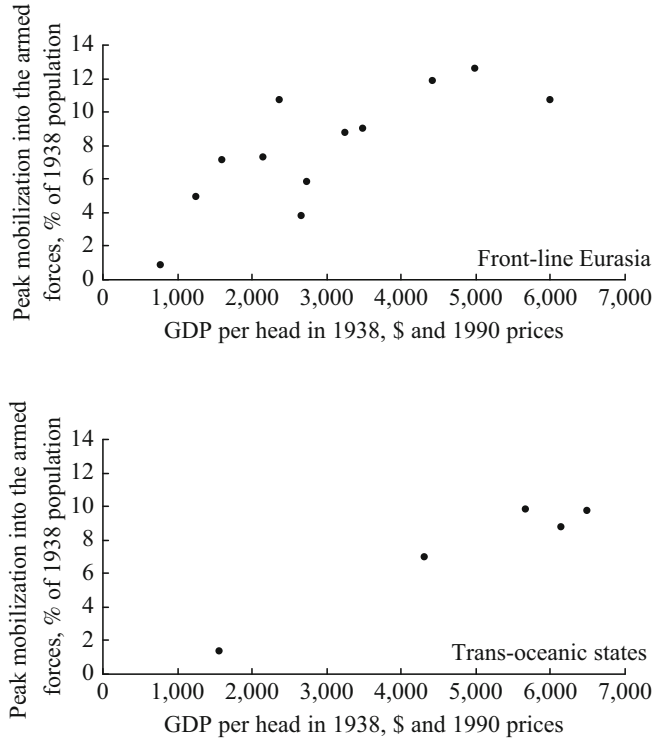
World War, but did not. The course of inward-looking industrialization that Stalin pursued between the wars does not appear to be a sufficient explanation. More important was Stalin's victory in the destructive struggle to collectivize farming, which ensured state control over wartime food supplies and prevented the peasants from seceding from the war effort (Gatrell and Harrison 1993). As a result the Soviet economy carried a disproportionately heavy economic burden in the Second World War without collapsing.

Finally, Fig. 6 shows military mobilization; again, the rich countries mobilised much higher proportions of their population into military uniform. The figure also shows the moderating effect of distance: when we control for pre-war incomes, the countries separated from the fighting by oceanic distances put fewer men into the fighting forces. But the effect of distance was less in the Second World War than in the First World War, suggesting that the interwar decline of transport costs had brought about a more truly global struggle.

Inter-Allied Cooperation

In terms of cooperation within the opposing coalitions, the Second World War saw a repeat of the First World War with some differences. In both wars the German-led coalition failed to achieve significant economic cooperation among the powers, each of which aimed primarily to exploit its own internal and colonial spaces. The Allies, in contrast, achieved fuller cooperation. During the first war, this involved pooling the industrial and commercial resources of Britain and America with the fighting strength of France, Italy, and Russia; the result was to permit the aggregate military power of the Allies to be produced more efficiently. The main instruments of pooling were war credits from America to Britain, France and Italy, and from France and Britain to Italy and Russia. The amount was not enough to keep Russia in the war to the end, but enough that post-war repayments significantly complicated post-war international finance and trade.

The second time round, inter-allied cooperation assumed a larger scale. The main form it took was the transfer of industrial goods – equipment (including vehicles), materials, fuels and processed foodstuffs – from the United States to Britain and from both countries to the Soviet Union. Although the US legislative framework called it 'Lend-Lease', the goods were actually



World Wars, Economics of, Fig. 6 Military mobilization in the Second World War: 17 countries (*Note*: observations are, from left to right: Front line Eurasia: China, Roumania, Bulgaria, USSR, Japan, Hungary, Greece, Italy, Finland, France, Germany and UK. Trans-Oceanic states: South Africa, Canada, Australia, USA and New Zealand. *Sources*: Harrison (1998, pp. 3–9, 14); Correlates of War data-set, version 2.1, at http://www.umich.edu/Bcowproj; Singer (1979,1980))

supplied free of financial charges, the aim being to promote the Allied partnership. Pooling of the resources counted in Tables 6 and 7 augmented their value, increased the Allied advantage, saved lives and resources, helped to prevent Soviet defeat, and brought forward the Allied victory.

Allied cooperation was not problem-free. The main issue was that, while it saved lives and brought forward victory, it did so asymmetrically. By keeping the Russians in the war, it saved primarily American and British lives, and the Russians felt this deeply. On the other hand, the victory that it brought forward was brought to Berlin by the Red Army, and was much more favourable to post-war Soviet power than would have been the case without western assistance – a source of wartime chagrin and post-war recriminations among the donors.

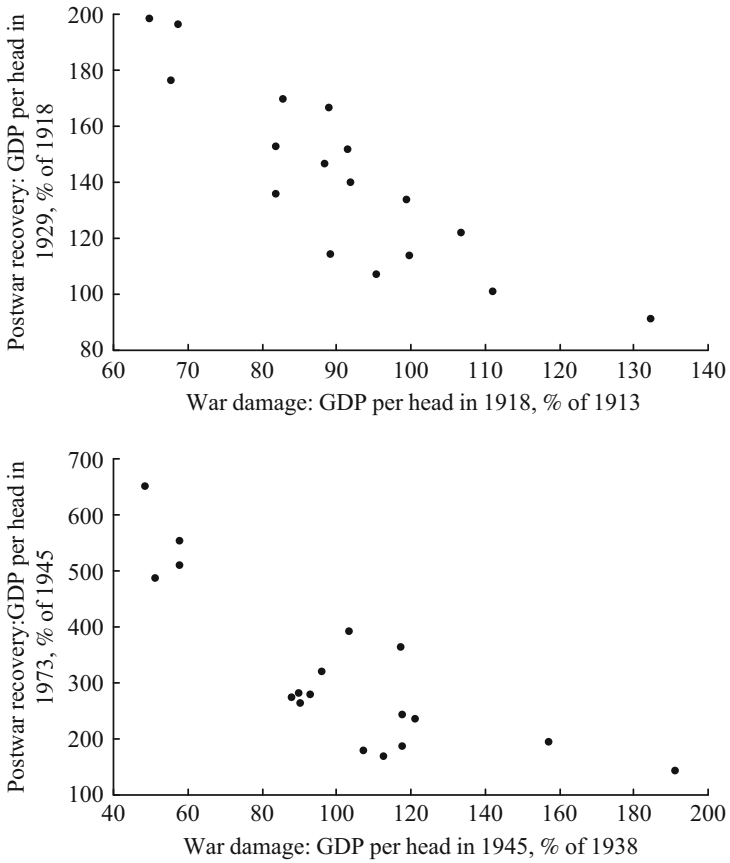
World Wars, Economics of, Table 9 War losses attributable to physical destruction (per cent of assets)

	Human assets	Physical assets	
		National wealth	Industry fixed assets
Allied powers			
USA	1	0	...
UK	1	5	...
USSR	18–19	25	...
Axis powers			
Germany	9	...	17
Italy	1	...	10
Japan	6	25	34

Source: Harrison (1998, p. 37)

War Losses

The Second World War was fought on a global scale but half a dozen countries saw most losses of wealth and population. Nearly all the 55 million



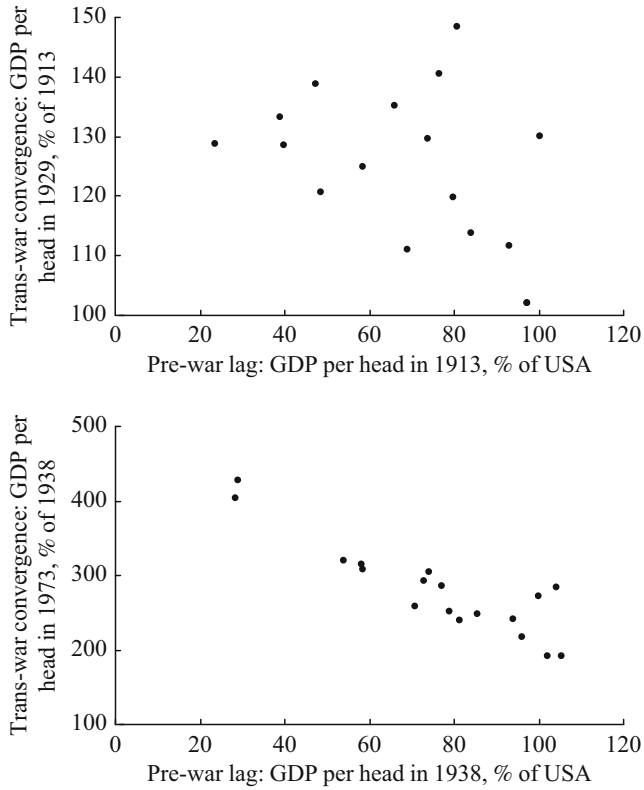
World Wars, Economics of, Fig. 7 Economic recovery following two world wars (*Note*: observations are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK and USA. *Source*: Maddison (2001))

premature deaths, for example, are accounted for between the USSR (25 million), China (10 million), Germany (6.5 million), Poland (5 million), Japan (2.4 million) and Yugoslavia (1.7 million). Table 9 summarizes the data for the great powers as percentages of prewar populations and assets.

The figures show the heavy loss of life in the Soviet Union, followed by German and Japan, and also the widespread destruction of property in the same countries. Everywhere, it seems, human capital was destroyed at a higher rate than physical capital. The survivors were endowed, therefore, with a ratio of physical to human capital that was advantageous by pre-war standards, provided that mismatches resulting from the wartime distribution of combat could be smoothed out. Table 9 takes no account of accelerated wartime investments in industry; in western Germany, for example, industrial capacity was added at a faster rate than bombing took it away, so that West German industry ended the war with a larger and newer stock of equipment than before (Abelshauser 1998, p. 168).

Economic Growth

Evidently, wartime economic mobilization tended to make the rich richer and the poor poorer. Thus, both wars tended to polarize the global distribution of income. It is of some interest, therefore, to examine whether postwar recovery and long-term economic growth succeeded in reversing this pattern.



World Wars, Economics of, Fig. 8 Economic convergence through two world wars (*Notes* and *Sources*: as for Fig. 7)

Figure 7 suggests that each war was followed by recovery and that those economies most damaged by the wartime experience recovered most rapidly. It takes 1929 as the benchmark date for measuring recovery from the First World War, and 1973 as the benchmark for recovery from the Second World War. It shows that, the more a country's average income fell during each war, the more it tended subsequently to rise. Thus, at least some of each war's negative effects were transitory.

A more complex picture emerges when we turn to long-term economic growth. To what extent did the post-war recovery return each country to a path of convergence on the global productivity frontier? Figure 8 suggests that after the First World War there was little or no convergence. Some countries that were already rich did much better after the war than some countries that were already poor. In contrast, the Second World War was followed by convergent economic growth. This suggests that the Allies designed a much better international environment for genuine convergence after 1945 than after 1918.

Conclusions

In this article we have shown that economics mattered, and we have shown how. Given time, resources won the two world wars. In mobilizing resources, the richer market economies had a significant advantage. It was more important to be rich than self-sufficient; probably, most pre-war efforts to protect jobs or diminish national dependence on trade in the name of strategic self-sufficiency were counterproductive. Poor economies, especially those with a large peasant population, tended to collapse under the stress of total war, although they tended to be less reliant on external trade. The main exception is the Soviet economy in the Second World War; its

exceptional resilience is best explained by its rulers' exceptional degree of control over the peasant farmers.

The pattern should not be overgeneralized. Broadberry and Harrison (2005) have suggested that the power of the relationship between economic and military performance is confined to a relatively short historical period. The era of 'total war' from 1914 to 1945 seems to have been unique. In both world wars the main combatants were able to devote more than half of their national income to the war effort. This is likely to have been impossible before 1914 because until then most people were too poor to be taxed at such rates; most economies had the bulk of their resources locked up in forms of subsistence agriculture that were resistant to mobilization; before mass literacy and the telegraph, typewriter and duplicator, commercial and government services were too inefficient to do much about it. In short, in earlier stages of global development total war could not be staged because too many people were required to labour in the fields and workshops just to feed and clothe the population, and it cost too much for government officials to count, tax and direct them into mass combat.

Since 1945 the economic factors in warfare may have lost significance again. This is because nuclear weapons can give devastating military force to any rich country however small, or any large country however poor, for a few billion dollars. Hence the marshalling of economic resources played a much more vital role in the outcome of the two world wars than was likely in any period before or after.

See Also

- ▶ Defence Economics
- ▶ War and Economics

Bibliography

Abelshauser, W. 1998. Germany: Guns, butter, and economics miracles. In *The economics of World War II: Six great powers in international comparison*, ed. M. Harrison. Cambridge: Cambridge University Press.

Abramovitz, M. 1986. Catching up, forging ahead and falling behind. *Journal of Economic History* 46: 385–406.

Bogart, E.L. 1920. *Direct and indirect costs of the great world war*. 2nd ed. New York: Oxford University Press.

Bowley, A.L. 1930. *Some economic consequences of the great war*. London: Thornton Butterworth.

Broadberry, S.N. 1998. How did the United States and Germany overtake Britain? A sectoral analysis of comparative productivity levels, 1870–1990. *Journal of Economic History* 58: 375–407.

Broadberry, S., and M. Harrison, ed. 2005. *The economics of World War I*. Cambridge: Cambridge University Press.

Broadberry, S.N., and P. Howlett. 1998. The United Kingdom: Victory at all costs. In *The Economics of World War II: Six great powers in international comparison*, ed. M. Harrison. Cambridge: Cambridge University Press.

Broadberry, S.N., and P. Howlett. 2005. The United Kingdom during World War I: Business as usual? In *The economics of World War I*, ed. S.N. Broadberry and M. Harrison. Cambridge: Cambridge University Press.

Carré, J.-J., P. Dubois, and E. Malinvaud. 1976. *French economic growth*. Oxford: Oxford University Press.

Ferguson, N. 1998. *The pity of war*. London: Allen Lane.

Gatrell, P. 2005. Poor Russia, poor show: Mobilising a backward economy for war, 1914–1917. In *The economics of World War I*, ed. S.N. Broadberry and M. Harrison. Cambridge: Cambridge University Press.

Gatrell, P., and M. Harrison. 1993. The Russian and Soviet economy in two world wars. *Economic History Review* 46: 425–452.

Goldsmith, R.W. 1946. The power of victory: Munitions output in World War II. *Military Affairs* 10(1): 69–80.

Hardach, G. 1987. *The First World War, 1914–1918*. Harmondsworth: Penguin.

Harrison, M., ed. 1998. *The economics of World War II: Six great powers in international comparison*. Cambridge: Cambridge University Press.

Harrison, M. 2005. The USSR and total war: Why didn't the Soviet economy collapse in 1942? In *A world at total war: Global conflict and the politics of destruction, 1939–1945*, ed. R. Chickering, S. Förster, and B. Greiner. Cambridge: Cambridge University Press.

Hautcoeur, P.-C. 2005. Was the great war a watershed? The economics of World War I in France. In *The economics of World War I*, ed. S.N. Broadberry and M. Harrison. Cambridge: Cambridge University Press.

Lloyd, E.M.H. 1924. *Experiments in state control*. Oxford: Oxford University Press.

Maddison, A. 2001. *The world economy: A millennial perspective*. Paris: OECD.

Milward, A.S. 1977. *War, economy and society, 1939–1945*. London: Allen Lane.

Milward, A.S. 1984. *The economic effects of the Two World Wars on Britain*. 2nd ed. London: Macmillan.

Offer, A. 1989. *The First World War: An agrarian interpretation*. Oxford: Clarendon Press.

Olson, M. 1963. *The economics of the wartime shortage: A history of British food supplies in the Napoleonic War and in World Wars I and II*. Durham: Duke University Press.
Ordway, F.I., and M.R. Sharpe. 1979. *The rocket team*. London: Heinemann.
Overy, R.J. 1995. *Why the allies won*. London: Pimlico.
Overy, R.J. 1998. Who really won the arms race? *The Times Literary Supplement.* 4–5.
Pamuk, S. 2005. The Ottoman economy in World War I. In *The economics of World War I*, ed. S.N. Broadberry and M. Harrison. Cambridge: Cambridge University Press.
Ránki, G. 1993. *The economics of the Second World War*. Vienna: Böhlau.
Ritschl, A. 2005. The pity of peace: Germany's economy at war, 1914–1918. In *The economics of World War I*, ed. S.N. Broadberry and M. Harrison. Cambridge: Cambridge University Press.
Schulze, M.-S. 2005. Austria-Hungary's economy in World War I. In *The economics of World War I*, ed. S.-N. Broadberry and M. Harrison. Cambridge: Cambridge University Press.
Sen, A.K. 1983. *Poverty and famines: An essay on entitlement and deprivation*. Oxford: Oxford University Press.
Singer, J.D. 1979/1980. *The correlates of war I: Research origins and rationale. II: Testing some realpolitik models*. New York: Free Press.
Urlanis, B. 1971. *Wars and population*. Moscow: Progress Publishers.
Villa, P. 1993. *Une analyse macroéconomique de l'économie française au XXe siècle*. Paris: CNRS.

Woytinsky, Wladimir Savelievich (1885–1960)

Peter Newman

Wladimir Woytinsky was born in St Petersburg, Russia, on 12 November 1885, the son of a professor of mathematics at a polytechnic college. Tutored at home until he was fourteen, at high school he was a brilliant student and in 1904 entered St Petersburg University to read law.

While still in high school he had written *Market and Prices*, which was published with a foreword by Tugan-Baranovsky (it gained another foreword, by Jacob Marschak, when it was translated and published four years after his death). As a subjectively original contribution to neoclassical economic theory its long technical discussion of demand theory and spatial competition made it respectable but unremarkable, but as a schoolboy's performance it was quite extraordinary.

By the time the book was published in 1906 'I turned the pages as though they had been written by a complete stranger' (1962, p. 9). The bloody riots of 1905 had shocked Woytinsky into revolutionary activity as a member – 'Sergei Petrov' – of the Bolshevik wing of the Social Democrats (S-D), though he seems never to have been a full marxist in his economics. He was in and out of Tsarist prisons for the next several years (including the notorious Ekaterinoslav Castle) and in the intervals continued his S-D activity, such as organizing public works for the unemployed in St Petersburg. Eighteen months after his last arrest he was sentenced in the summer of 1909 to four years of forced labour, to be followed by deportation for life to Siberia.

Most of his time in Siberia was spent out of jail, and in 1916 he met and married an energetic young woman, Emma Shadkhan, who became his devoted assistant and co-author for the rest of his life. When the Revolution broke out in 1917 he returned to Petrograd and began to work for the Menshevik wing of the S-D, influenced in particular by Tseretelli. He spent some time as an editor of Izvestia and then became the Provisional Government's Commissar on the Northern Front. After the fall of that Government in November he was arrested by Trotsky himself but released early in 1918. In disguise, he and his wife fled with Tseretelli to Georgia, where Woytinsky was asked to edit the political newspaper *Bor'ba*.

In August 1920 he became economic adviser to a Georgian delegation to the countries of Western Europe, and he travelled with them until soon after Georgia was absorbed into the USSR in 1921. He settled in Germany in 1922, began work as a private economic journalist, and with Ladislaus von Bortkiewicz as editorial adviser, produced the massive *Die Welt in Zahlen* (The World in Figures), which established his reputation in Germany. In 1928 he was asked to direct the statistical department of the General Federation of Labour Unions and to act as its economic consultant, which brought him into contact with many German Social Democrats. With the onset

of the Great Depression his advocacy of public works financed by budget deficits received approval from economists such as Gerhard Colm but met with considerable opposition in the party, especially from Rudolf Hilferding.

Hitler's mass arrest of union leaders in 1933 convinced Woytinsky to leave Germany. 'I was a foreigner, a Russian Jew active in the German labour movement' (1961, p. 477). After brief stops in Switzerland and France he joined the International Labour Office (I.L.O) in Geneva for a year and then, in 1935, the couple left for the United States, where he worked for the Central Statistical Board (1935–6), the Committee on Social Security of the Social Science Research Council (1936–41), and finally the Social Security Administration itself (1941–7). His career ended as a research director for the Twentieth Century Fund, combined with a research professorship at Johns Hopkins University. He died on 10 June 1960, in Washington, DC.

Woytinsky's main professional output consisted of massive compendia of economic statistics produced mostly by him and his wife working alone, a form of work and publication that for better or worse is now quite out of fashion in economics. His other published studies were mostly concerned with current economic and social policy and their interest was largely ephemeral. So his work seems unlikely to be remembered; but his life exemplifies what can happen in this terrible century to an intellectual with a social conscience.

Selected Works

1906. *Rynok i tsieny*. St. Petersburg, with a foreword by M.I. Tugan-Baranowsky. Translated by Emma Woytinsky as *Market and prices. Theory of consumption, market and market prices*, with an additional foreword by Jacob Marschak. New York: Augustus M. Kelley, 1964. (Curiously, this translation nowhere contains the Russian title of the original book, either transliterated or in cyrillic.)

1925–8. *Die welt in Zahlen*, 7 vols. Berlin: R. Mosse.

1926. *Die vereinigten staaten von Europa*. Berlin: J.H.W. Dietz nachf.

1935. *Three sources of unemployment*. Geneva: ILO.

1936. *The social consequences of the economic depression*. Geneva: I.L.O.

1939. *Seasonal variations in employment in the United States*. Washington, DC: Social Science Research Council.

1942. *Three aspects of labor market dynamics*. Washington, DC: Social Science Research Council.

1943. *Earnings and social security in the United States*. Washington, DC: Social Science Research Council.

1953. (With associates.) *Employment and wages in the United States*. New York: Twentieth Century Fund.

1955. (With E.S. Woytinsky.) *World commerce and governments; Trends and outlook*. New York: Twentieth Century Fund.

1962. *Stormy passage*. Introduction by A.A. Berle. New York: Vanguard Press.

Bibliography

Woytinsky, E.S. (ed.). 1962. *So much alive; The life and work of Wladimir S. Woytinsky*. New York: Vanguard Press.