6

THE THEORETICAL CRITIQUE OF WALRASIAN WELFARE ECONOMICS

It is the first welfare theorem [asserting the efficiency of competitive markets] that provides the intellectual foundation for our belief in market economies. Like any theorem its conclusions depend on the validity of its assumptions. A closer look at those assumptions, however, suggests that the theorem is of little relevance to modern industrial economies.

—*Joseph Stiglitz,* Whither Socialism? *(Cambridge, MA: MIT Press, 1994), 28*

THE IMPORTANCE OF THE FUNDAMENTAL THEOREMS

The foundation for Walrasian economic theory is the two welfare theorems discussed in chapters 4 and 5. The First Fundamental Theorem states that if all individuals and firms are selfish price takers, then a competitive equilibrium is Pareto efficient. The Second Fundamental Theorem states that if all individuals and producers are selfish price takers, then almost any Pareto optimal equilibrium can be supported via the competitive mechanism, provided appropriate lump-sum taxes and transfers are imposed on individuals and firms.

These two theorems have been the backbone of economic policy in the decades since World War II. Lockwood (1987, 811) writes of the second theorem: "It is no exaggeration to say that the entire modern microeconomic theory of government policy intervention in the economy (including cost-benefit

This chapter was originally published in a somewhat different form in John M. Gowdy, "The Revolution in Welfare Economics and Its Implications for Environmental Valuation and Policy," *Land Economics* 80(2) (2004), 239–257.

analysis) is predicated on this idea." Likewise, Fisher (1983) writes: "The central theorems of welfare economics (i.e. the first and second fundamental theorems) may be the single most important set of ideas that economists have to convey to lay people."

The second theorem implies that if a move from one particular state of the economy to another is judged to be desirable, this move may be achieved through transferring resources from one person (or one activity) to another. Referring back to Chapter 1, any point on the contract curve in Figure 1.3 can be reached by changing the initial distribution of the goods X and Y among the two consumers. The rationale for moving from one state of the economy to another is the Kaldor–Hicks criterion, that is, identifying a potential Pareto improvement (PPI). If the magnitude of the gains from moving from one state of the economy to another is greater than the magnitude of the losses, then social welfare is increased by making the move even if no actual compensation is made. Stavins, Wagner, and Wagner (2002, 5) write about the potential Pareto improvement: "This is the fundamental foundation—the normative justification—for employing benefit-costs analysis, that is, for searching for policies that maximize the positive differences between benefits and costs."

Establishing economic policies using cost-benefit analysis to identify PPIs is one of the central concerns of contemporary economics. A PPI is fundamentally different from the notion of Pareto efficiency that simply says that an efficient state is one in which any change will make at least one person worse

off. A PPI is a change that helps one person and harms another. PPIs are identified using the measures of consumer surplus discussed in Chapter 5.

Two problems mar the welfare-based cost-benefit approach to economic policy. The first is the intractable theoretical difficulty of determining PPIs using the Kaldor–Hicks criterion. This is discussed in this chapter. The second problem is the characterization of human behavior to fit the restrictive assumptions of consumer choice theory. This is discussed in Chapter 7.

What follows is *not* a critique of all attempts to quantify the benefits and costs of moving from one state of the economy to another. Problems arise to the extent that estimates of costs and benefits are shoehorned into the narrow framework of Walrasian theory. Many economists do not appreciate the theoretical difficulties involved in estimating welfare changes. And certainly most non-economists are unaware of the leap of faith required to move from estimating costs and benefits to calculating a potential Pareto improvement.

A Summary of the Walrasian System

1. The theoretical foundation for the policy recommendations of Walrasian economics is the first two theorems of welfare economics.
2. The First Fundamental Theorem justifies relying on a competitive economy to ensure the social good.
3. The Second Fundamental Theorem justifies interventionist policies but only to create the conditions for competitive markets.
4. The system of equations supporting these theorems depends critically on the assumptions about consumer behavior (*Homo economicus*) and the characterization of markets and technology (perfect competition).
5. In this system, policy recommendations are based on the ability to identify efficiency gains (potential Pareto improvements).
6. The ultimate goal of this system is to create a *positive* science of economics, one that can provide policy recommendations without making value judgments, that is, without making interpersonal comparisons of utility.

THEORETICAL INTRACTABILITIES WITH IDENTIFYING POTENTIAL PARETO IMPROVEMENTS

The Kaldor–Hicks criterion seems straightforward. If one person values his or her gains from an economic change more than a second person values his or her losses, then potential total welfare increases, and this represents a potential Pareto improvement. Such an economic change is justified on efficiency grounds even if no actual compensation is paid. In general, economists follow Kaldor's view that economic policy recommendations should be determined by efficiency; distribution is a problem for politicians. Undermining the argument for separating efficiency and distribution is more than sixty years of theoretical work demonstrating that PPIs cannot be identified by comparing individual welfare changes. The goal of economists for most of the last century was to make economics a "positive" science, and the focus on efficiency was supposed to accomplish this. Finding efficiency improvements was supposed to allow economists to avoid interpersonal comparisons of utility. This proved to be an impossible task. The problem in applying the potential Pareto principle is that we are drawing **general equilibrium** conclusions from **partial equilibrium** situations. Partial equilibrium analysis assumes that changes in a particular market can be analyzed independently from all other markets. This is a useful assumption in some practical policy applications of economic theory but is not valid if one is trying to establish the conditions for a general economic optimum. This point is illustrated by the following paradoxes in welfare theory. These paradoxes are particularly debilitating to Walrasian theory because they show that even if we accept all the basic assumptions about economic man and perfectly competitive markets, the theory is still internally inconsistent.

The Cycling Paradox

The PPI criterion was supposed to allow economists to make policy recommendations regarding any two points on different utility possibilities curves such as those shown in Figure 3.4 in Chapter 3. Theoretical difficulties were raised immediately after Kaldor and Hicks proposed in 1939 that identifying PPIs should be the goal of economics. In 1941, Tibor Scitovsky demonstrated

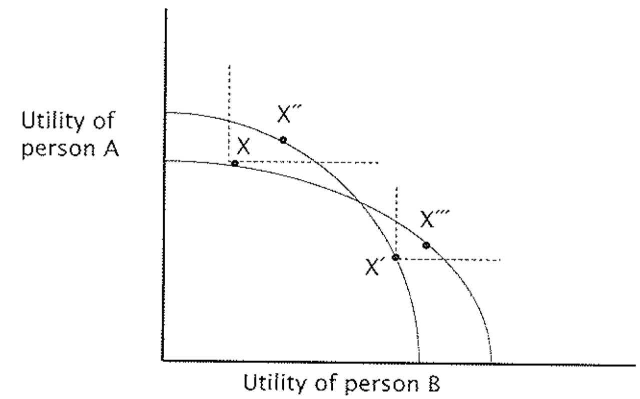


Figure 6.1. The cycling paradox (adapted from Varian 1992, 406)

that if a movement from one point to another in utility space can be shown to be Pareto improving according to the Kaldor–Hicks criterion, then it may also be shown that a movement back to the original point is also Pareto improving. This is sometimes called the cycling paradox. Referring to Figure 6.1, using the PPI criterion a movement from point X to point X′ should be made because from X′ it is possible to move to X″ where both consumers are better off compared with the original point X. It is also true, however, that a movement from X′ to X is justified because from X it is possible to move to point X‴ where both consumers are better off compared with the starting point X′.

To eliminate this cycling problem, Scitovsky proposed a double criterion for a potential Pareto improvement. It must be shown that the gainers from a change can compensate the losers so they will agree to the change (Kaldor criterion) and that it is not possible for the losers to bribe the gainers not to make the move (Hicks criterion). But Gorman (1955) showed that the Scitovsky criterion violates the assumption of transitivity, which as we saw in Chapter 1 is a necessary condition for consistency in consumer choice.

The "Sticking" Paradox

The PPI was one attempt to broaden the Pareto criterion without making interpersonal comparisons of utility. Another attempt, discussed in Chapter 3,

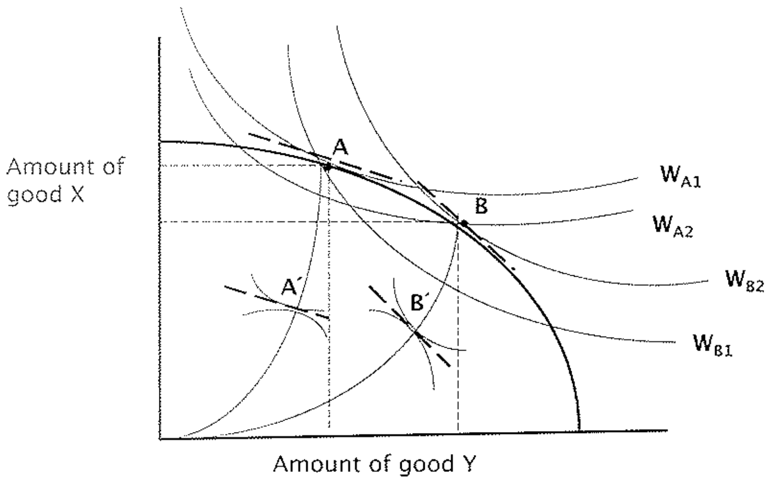


Figure 6.2. A production possibilities frontier with two social welfare optima

was constructing a social welfare function (SWF) to choose a point on a grand utilities possibilities frontier or on a production possibilities frontier (PPF). Figure 6.2 shows that using a SWF might "stick" the choice at either A or B, depending on which was the starting point. If the initial Pareto equilibrium is at B with associated social welfare functions W_{B1} and W_{B2}, point B would be preferred to point A because it is on a higher social welfare curve. However, if the starting point is A with its associated social welfare curves W_{A1} and W_{A2}, then point A would be preferred to point B.

The problem is that any change in initial distribution of goods (income) means a change in the reference points that determine Pareto optimality. The points A and B on the production possibilities frontier are associated with points A′ and B′ within an Edgeworth box for each amount of goods X and Y. Each utility possibilities frontier in Figure 6.2 can be derived from one of the two contract curves for consumption. As we saw in Chapter 3, the necessary condition for general Pareto optimality is that the slope of the production possibilities frontier, the rate of product transformation of Y into X ($RPT_{Y\ for\ X}$), is equal to the common marginal rates of substitution Y for X ($MRS_{Y\ for\ X}$) in consumption for each person. These slopes will be different at different points

along the PPF, meaning that in competitive equilibrium, the price ratios for X and Y will be different at points A and B. This general interdependence of welfare distributions and relative prices means that we cannot make general equilibrium statements comparing points on the PPF.

The Boadway Paradox

Comparisons of the relative efficiencies of different economic situations depend on identifying gains in consumer surplus (or compensating variations and equivalent variations—CVs and EVs) as discussed in Chapter 5. According to standard cost-benefit practice when comparing different projects or policies, judging which one is superior is a matter of finding the one with the largest net gain. In the 1970s, Robin Boadway (1974) demonstrated that when comparing alternatives, the one with the highest net gain is not necessarily the "best" one as judged by the Kaldor–Hicks compensation test. This is referred to as the Boadway paradox, and it also arises from the fact that estimates of income-compensated variations—welfare gains at constant prices—are partial equilibrium measures. If relative prices change with a redistribution of income, as they almost certainly would in a general equilibrium system, then such estimates are misleading measures of potential welfare gains. These measures coincide with general equilibrium measures only if there is a single market-clearing price ratio at every point on the contract curve, a condition that holds only if preferences are identical and homothetic.

Homothetic Utility Functions

The assumption of "homothenicity" is critical to the Walrasian system and is worth going into in some detail. A homothetic utility function means that the utility maximizing composition of consumption goods depends only on the relative prices of the goods, not income, as shown in Figure 6.3.

A line drawn from the origin (the expansion path of consumption as income increases) will cross the tangency points of the budget lines and indifference curves where these have the same slopes for all levels of income. This means that the marginal rate of substitution between the two goods does not change as income increases. Put another way, the equilibrium (utility maximizing) proportion of all goods consumed remains exactly the same as

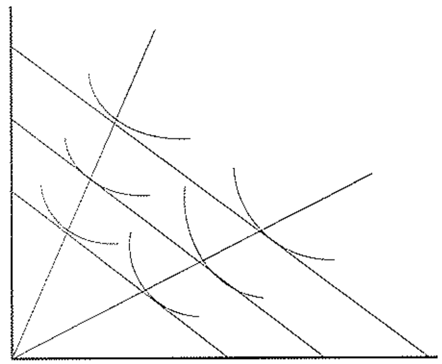


Figure 6.3. Expansion paths for homothetic utility functions

income increases. This is an extremely restrictive and unrealistic assumption. The relative amounts of goods X and Y chosen in Figure 6.3 depend solely on their relative prices. Homothetic utility functions exhibit income elasticities of demand equal to one for all goods for all levels of income. For the whole economy this means that consumers with different incomes but facing the same set of relative prices (as under perfect competition) will demand goods in the same proportions.

Homothenicity is an extremely unrealistic assumption, even in the context of neoclassical economics, but it is necessary to save the Walrasian system and avoid the above paradoxes. Remember that the critical assumption underlying the mathematical properties of consumption and production theory is that there is no interaction between agents. With the homothenicity assumption, everyone has the same preferences and identical marginal utilities of income. An increase in income for the richest person on the planet and the poorest would result in the same proportionate increase in the commodities they consume. People would have identical tastes (utility functions) and there would be no reason to trade goods. The necessity of the homothenicity assumption is one reason why economists resort to the notion of a "representative agent." This is the common theoretical practice of using a single

individual to serve as a proxy for all consumers or all firms. The assumption of homothetic preferences, and identical preferences, is obviously a gross violation of reality that must have profound effects on the results of empirical analyses. But it is necessary given the mathematical requirements of Walrasian economics.

A final irony here is worth mentioning. Recall that the goal of **positive economics** is to avoid interpersonal comparisons of utility. But the assumption that utility functions are homogeneous and that the marginal utility of income is the same for all consumers cannot be made without making interpersonal comparisons of utility.

More Problems with PPIs

Numerous other theoretical dilemmas with the PPI approach have been identified. Kjell Brekke (1997) showed that the choice of a numeraire matters when the marginal rates of substitution differ among consumers. Samuelson (1950) showed that it is not certain that group A is better off than group B even if group A has more of everything. Again, a basic problem for welfare economics is that the axioms of consumer choice refer to a single individual and they break down in the case of two or more persons. In the case of two or more persons, even within the narrow framework of neoclassical welfare theory, it cannot even be proved that more is preferred to less—perhaps *the* basic assumption of modern economics (Bromley 1990).

The upshot of these results for welfare economics is that the Kaldor–Hicks PPI rationale for comparing two states of the economy has some fundamental problems that makes it unacceptable as a theoretical foundation for analyzing the costs and benefits of economic policies. There is no theoretically justifiable way to make welfare judgments without invoking value judgments and interpersonal comparisons of utility, yet this is not permissible under the stringent requirements of neoclassical welfare economics. Chipman and Moore (1978, 581) summarized the outcome of discussions about the Kaldor–Hicks–Samuelson–Scitovsky new welfare economics: "After 35 years of technical discussions, we are forced to come back to Robbins' 1932 position. We cannot make policy recommendations except on the basis of value judgments, and these value judgments should be made explicit." This position is even more secure after another twenty-five years of theoretical discussions.

Efficiency, Output Mix, and Social Welfare

As shown in Figure 6.4, it is quite possible that an increase in "efficiency" can reduce social welfare if the output mix changes even if total output increases. Suppose a technological improvement, indicated by an outward shift in the production possibilities frontier, moves the economy from point B to point A on a higher PPF. If we assume that total welfare increases with total consumption, this move should be made under the Kaldor–Hicks test because total consumption (output) goes up. But social welfare declines as indicated by the move from the social welfare function W_3 to a lower social welfare function W_2. In economic arguments for growth, the separability principle is extended to say that the output mix is a political and not an economic problem. It is claimed that efficiency is a "positive" goal, and the question of the mix of goods and productive inputs involves "normative" judgments. An increase in efficiency is a good thing because it is theoretically possible for political authorities to redistribute the efficiency gains so that we could end up at point C on the higher social welfare curve W_4. In practice, the drive for efficiency almost always overrides equity.

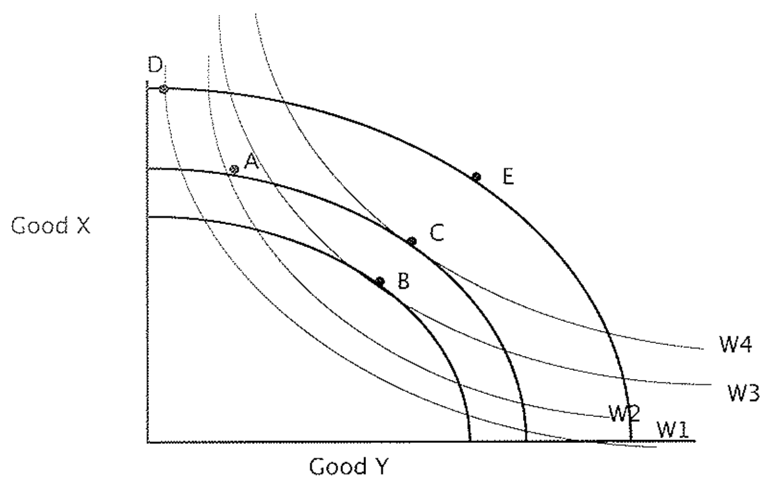


Figure 6.4. Efficiency and social welfare: Efficiency trumps equity

Applying the efficiency rule would dictate a move from point A not to point C but rather to point D on a higher PPF but an even lower social welfare curve W_1. Output mix as well as income distribution, both ignored by the PPI criterion, are essential elements of social welfare.

WHY IS THE THEORETICAL CRITIQUE IMPORTANT?

A common defense of Walrasian microeconomics goes something like this: "Of course we know all this. No one really believes that the **real economy** is characterized by the general equilibrium model we use to describe it. But it is good enough." But this model is *the* intellectual foundation for free market policies. If it can be shown that the mathematical logic of the model does not support the fundamental theorems, then there is a crisis of confidence problem for microeconomic theory. If we are no longer sure that the theoretical system generates a Pareto efficient outcome, what does this say about actual markets? Without the Walrasian core to appeal to, economic policies have to be argued case by case with hard empirical evidence about actual costs and benefits. It is no longer enough to set markets in motion and say that a socially optimal outcome will *automatically* be the result.

Exercise

Almost every introductory economic textbook has a discussion about the inefficiency of a minimum wage for workers.

1. Find such a discussion and list the underlying assumptions.
2. Go to an economic literature search engine (or simply Google "minimum wage") and classify the articles (or blogs) according to whether they are for or against a minimum wage.
3. Categorize the groups as to whether they are based on empirical evidence or on Walrasian theory.

IS THE GOAL OF ECONOMIC EFFICIENCY "SCIENTIFIC" OR IS IT AN "IDEOLOGY"?

The policy effect of the development of the various currents and concepts discussed in chapters 1–5 was to focus economics on the concept of efficiency.

In the decades after World War II the job of policy-oriented economists became one of identifying market imperfections, or, in the language of some, finding "money on the table." If resources could be diverted from one use to another and increase total welfare (that is, total economic output), then everyone could (potentially) be better off. This was the job of cost-benefit analysis, and it seemed to finally cast economics as a positive, value-free science. But, as we have seen in this chapter, the theoretical justification for identifying efficiency improvements, the potential Pareto improvement, foundered on theoretical intractabilities. Economic efficiency can no longer claim to be an objective policy criterion.

Daniel Bromley (1990) argues that the claim for "efficiency" as a policy goal is based on ideology, not science. He uses the word "ideology" not in a pejorative sense but rather in the sense of "a shared system of meaning and comprehension." Efficiency is not an objective goal but rather an opinion based on the system of beliefs embodied in the worldview of the Walrasian system. Bromley points out that a (probably unintended) consequence of Arrow's theorem was to divert public policy from democratic deliberation to the market. The first sentence in Arrow's book *Social Choice and Individual Values* (1951) states: "In a capitalist democracy there are essentially two methods by which social choices can be made: voting typically used to make 'political' decisions, and the market mechanism, typically used to make 'economic' decisions" (quoted in Bromley 1990, 92). The competitive market, with all the built-in assumptions that term implies, was a reliable mechanism to "scientifically" allocate resources to their most efficient uses without resorting to the messy (and "inefficient") process of democratic debate (sometimes called "the decision-making approach"). But as we have seen above, the identification of efficiency improvements (PPIs) not only is based on dubious behavioral assumptions, but it is also plagued by internal inconsistencies in the Walrasian system.

SUMMARY

In spite of mounting empirical evidence and a large body of theory demonstrating the logical inconsistencies and empirical shortcomings of Walrasian welfare economics, this framework continues to dominate economic textbooks. However, leading theorists, including many recent Nobel Prize winners in economics, have all but abandoned that framework. Judging from the contents of the leading economics journals, day-to-day work by applied economists is curiously disconnected from current work in mainstream economic theory. A time lag between theoretical frontiers and everyday practice is normal in any science, but its consequences are severe in the case of economic valuation. Current U.S. policies on climate change and biodiversity preservation, for example, rely heavily on welfare economic models whose legitimacy depends crucially on questionable theoretical formulations and on assumptions known to be wildly at odds with actual human behavior. Particularly problematic is the use of the concept of a potential Pareto improvement (PPI) as one of the major economic tools for evaluating alternative economic policies.

In a way Arrow's impossibility theorem summarizes the paradoxes discussed above. There is no way to aggregate the preferences of self-regarding individuals. The fatal flaw in Walrasian theory is that it cannot handle interdependencies among economic actors. Feldman (1987, 894) summarizes the uncomfortable state of neoclassical welfare economics after Arrow's theorem:

> Where does welfare economics stand today? The First and Second Theorems are encouraging results that suggest the market mechanism has great virtue: competitive equilibrium and Pareto optimality are firmly bound. But measuring the size of the economic pie, or judging among divisions of it, leads to paradoxes and impossibilities summarized by the Third Theorem. And this is a tragedy. We feel we know, like Adam Smith knew, which policies would increase the wealth of nations. But because of all our theoretic goblins, we can no longer prove it.

Perhaps the most important lesson for the theoretical literature in welfare economics is that we cannot do welfare economics without making interpersonal comparisons of utility. Many of the paradoxes and impossibilities embedded in Walrasian welfare economics arise from the attempt to avoid interpersonal comparisons of utility. People are, in fact, other regarding, and we cannot accurately portray economic behavior without accepting this fact. This observation dovetails with empirical findings from game theory and behavioral economics. Preferences are socially constructed and behavior is other regarding. This is the subject of the next chapter.

APPENDIX

Arrow's Impossibility Theorem

A formal (short version) definition of the Third Fundamental Theorem of Welfare Economics from the *New Palgrave Dictionary of Economics* is: "There is no Arrow Social Welfare function that satisfies the conditions of universality, Pareto consistency, neutrality-independence-monotonicity, and non-dictatorship." Arrow's theorem is sometimes called the **paradox of voting** and it can be illustrated as follows. Suppose we have three individuals (A, B, and C) with the following preferences for states of the economy (x, y, and z).

	Rank		
Person	1st	2nd	3rd
A	x	y	z
B	y	z	x
C	z	x	y

Suppose we do the "voting" in two stages. First we choose between x and y, and x is preferred to y by 2:1. Next we choose between y and z, and y is preferred to z by 2:1. This gives us the ranking in order of preference: x, y, z. But suppose we begin by choosing between x and z. In this case z is preferred to x by 2:1. Then we choose between x and y, and x is preferred 2:1. This gives us a different ranking: y, z, x.

The Microfoundations of Macroeconomics

The rise of Walrasian economics as a description of consumer and firm behavior combined with the notion of general equilibrium led naturally to using the same barter framework to describe entire economies. The model we first saw in Chapter 1 of individuals sitting around a table directly trading CDs became the preferred way to examine activity in regions, nations, and even the world economy. The concerns of classical economics—the distribution of income among economic classes or the long-term availability of natural resources, for example—were forgotten altogether or shoehorned into the Walrasian constrained optimization model. The utility function we saw in Chapter 1 became an aggregate demand function and production was analyzed in terms of aggregate supply. The applicability of marginal analysis to the macroeconomy was largely unexamined. Consider the idea of the "marginal product of labor" for the U.S. economy. What does it mean to examine the effect of adding a unit of labor to the whole economy, keeping everything else unchanged?

We can see how the barter economy model operates at the macro level by examining the **quantity equation,** the starting point for monetary economics, a school of thought popularized by Milton Friedman in the 1970s and 1980s. The quantity equation is:

$$MV \equiv PQ \qquad (6.1)$$

The left side of the equation shows the amount of money spent in a year. M is the money supply and V is velocity, or the number of times a unit of money (a dollar or a Euro) is spent during a particular time period, usually one year. So MV is the total amount of money consumers spend on goods and services in a year. The right side of the equation, PQ shows the value of goods sold in year. P is price of goods and Q is the physical quantity of goods. The equation is an identity because, obviously, the monetary value of goods sold must equal the monetary value of goods bought.

Monetarists made some strong assumptions about the quantity equation based on Walrasian economics. First, they assumed that the **velocity of money** was constant, that is, the proportions of money consumers decide to spend or save does not change appreciably over time. Second, they assumed that the real economy (the physical production of goods and services) is unaffected by the money supply. This is the no-money-illusion assumption imported into macroeconomics. If these assumptions are true, then the only effect of a change in the money supply, M, is to change the price level, P. Two policy implications of the quantity equation, and these assumptions, are (1) monetary policy (changing the money supply) can have no effect on the real economic activity, and (2) inflation is caused only by increases in the money supply (too many dollars chasing too few goods).

The monetarists used the quantity equation to argue for a "hands-off" economic policy on the part of governments. The only role of the monetary authority should be to keep the money supply growing at the same (constant) rate as real output, Q—so there would be no inflationary tendency. **Monetarism** has fallen out of favor because, among other things, it has been shown that neither basic assumption of the quantity equation holds—V is not constant and the real economy, Q, is affected by changes in the money supply, M. But the quantity equation (and monetary theory) is a good example of how macroeconomic theory and policies arise directly from the model of a simple barter economy and the assumptions economists make to jump from there to a market economy with money and prices.

GLOSSARY

Boadway paradox—When comparing alternatives, the one with the highest net gain is not necessarily the "best" one as judged by the Kaldor–Hicks compensation test.

Cycling paradox—If a movement from one point to another in utility space can be shown to be Pareto improving according to the Kaldor–Hicks criterion, then it may also be shown that a movement back to the original point is also Pareto improving.

General equilibrium—Economic equilibrium in all interrelated markets in the economy.

Hicksian income—The amount of economic product that is left for consumption after all capital stocks have been maintained intact. Some economists refer to this as "wealth" or "wealth per capita." It is GNP adjusted for externalities and capital (including "natural capital") depreciation.

Homothetic utility functions—A homothetic utility function implies that along the expansion path of consumption, indifference curves will have the same slopes. The marginal rate of substitution between two goods does not change as income increases. The utility maximizing proportion of all goods consumed remains exactly the same as income increases.

Ideology of efficiency—A term coined by Daniel Bromley to describe the mindset of economists who believe that the scope of economics is to identify potential Pareto improvements.

Microfoundations—The application of Walrasian theories of the consumer and firm to the analysis of the macroeconomy.

Monetarism—A school of economic thought, led by the late Milton Friedman, that advocates minimal government involvement in economic life. Monetarists argue that neither fiscal nor monetary policy is effective in stabilizing the economy or encouraging economic growth.

Paradox of voting—Arrow's impossibility theorem.

Partial equilibrium—Part of the economy is in equilibrium without reference to the rest of the economy. All prices and quantities outside that partial market are assumed to be constant and unaffected by changes within the partial market.

Positive economics—The view that economics should be about "what is" rather than "what ought to be." This is the rationale for trying to avoid interpersonal comparisons of utility.

Quantity equation—An equation for the macroeconomy showing the identity of the amount of money spent by consumers and the amount money received by producers (sellers) of those goods and services. It is written as $MV \equiv PQ$ and is the basic equation of monetary economics.

Real economy—The physical production of goods and services without reference to prices or money.

Velocity of money—The number of times a unit of money (a dollar or a Euro) is spent during a particular time period, usually one year.

REFERENCES AND FURTHER READING

Overviews of Welfare Theory

Bowles, S., and H. Gintis. 2000. Walrasian economics in retrospect. *Quarterly Journal of Economics* 115, 1411–1439.

Chipman, J., and J. Moore. 1978. The new welfare economics 1939–1974. *International Economic Review* 19, 547–584.

Feldman, A. 1987. Welfare economics. In *The New Palgrave Dictionary of Economics*, vol. 4, ed. J. Eatwell, M. Milgate, and P. Newman. London and New York: Macmillan, 889–895.

Fisher, F. 1983. *Disequilibrium Foundations of Equilibrium Economics*. Cambridge: Cambridge University Press.

Lockwood, B. 1987. Pareto efficiency. In *The New Palgrave Dictionary of Economics*, vol. 3, ed. J. Eatwell, M. Milgate, and P. Newman. London and New York: Macmillan, 811–813.

Varian, H. 1992. *Microeconomic Analysis*, 3rd ed. New York: Norton.

Early Critiques of Welfare Theory

Gorman, W. M. 1955. The intransitivity of certain criteria used in welfare economics. *Oxford Economic Papers*, new series, 7, 25–35.

Samuelson, P. 1950. Evaluation of real national income. *Oxford Economic Papers*, new series, 2, 1–29.

Scitovsky, T. 1941. A note on welfare propositions in economics. *Review of Economic Studies* 9, 77–88.

More Recent Theoretical Critiques

Albert, M., and R. Hahnel. 1990. *A Quiet Revolution in Welfare Economics*. Princeton, NJ: Princeton University Press.

Bromley, D. 1990. The ideology of efficiency: Searching for a theory of policy analysis. *Journal of Environmental Economics and Management* 19, 86–107.

Gowdy, J. 2004. The revolution in welfare economics and its implications for environmental valuation. *Land Economics* 80, 239–257.

Koning, N., and R. Jongeneel. 1997. Neo-Paretian welfare economics: Misconceptions and abuses. Wageningen Economic Papers, 05-97, Wageningen University, The Netherlands.

Stiglitz, J., 1994. *Whither Socialism?* Cambridge, MA: MIT Press.

van den Bergh, J. C. M., and J. Gowdy. 2003. The microfoundations of macroeconomics: An evolutionary perspective. *Cambridge Journal of Economics* 27, 65–84.

Social Choice Theory and Cost-Benefit Analysis

Arrow, K. 1951. *Social Choice and Individual Values*. New Haven, CT: Yale University Press.

Boadway, R. 1974. The welfare foundations of cost-benefit analysis. *Economic Journal* 47, 926–939.

Brekke, Kjell. 1997. The numeraire matters in cost-benefit analysis. *Journal of Public Economics* 64, 117–123.

Bryant, W. 1994. Misinterpretations of the second fundamental theorem of welfare economics: Barriers to better economic education. *Journal of Economic Education*, Winter, 75–80.

Chipman, J. 1974. The welfare ranking of Pareto distributions. *Journal of Economic Theory* 9, 275–282.

Chipman, J., and J. Moore. 1976. Why an increase in GNP need not imply an improvement in potential welfare. *Kyklos* 29, 391–418.

Stavins, R., A. Wagner, and G. Wagner. 2002. Interpreting sustainability in economic terms: Dynamic efficiency plus intergenerational equity. Discussion paper 02-29, August. Washington, DC: Resources for the Future.

Suzumura, K. 1999. Paretian welfare judgements and Bergsonian social choice. *Economic Journal* 109, 204–221.

THE BEHAVIORAL CRITIQUE OF WALRASIAN WELFARE ECONOMICS

Utility cannot be divorced from emotion, and emotions are triggered by changes. A theory of choice that completely ignores feelings such as the pain of losses and the regret of mistakes is not only descriptively unrealistic, it also leads to prescriptions that do not maximize the utility of outcomes as they are actually experienced.

—*Daniel Kahneman, "Maps of Bounded Rationality: Psychology for Behavioral Economics,"* American Economic Review *93(5) (2003), 1457*

Formidable criticisms of the behavioral assumptions of economic theory have been made for a century or more. Thorstein Veblen wrote this eloquent paragraph about economic man in 1898.

> The hedonistic conception of man is that of a lightning calculator of pleasures and pain, who oscillates like a homogeneous globule of desire of happiness under the impulse of stimuli that shift him about the area, but leave him intact. He has neither antecedent nor consequence. He is an isolated, definite human datum in stable equilibrium except for the buffets of the impinging forces that displace him in one direction or another. Self-imposed in elemental space, he spins symmetrically about his own spiritual axis until the parallelogram of forces bears down upon him, whereupon he follows the line of the resultant. When the force of the impact is spent, he comes to rest, a self-contained globule of desire as before. (Veblen 1898, 389–391)

Criticisms like Veblen's have had little effect on mainstream economics partly because they can be dismissed as "unscientific." As long as the arguments about

This chapter was originally published in a somewhat different form in John M. Gowdy, "Behavioral Economics and Climate Change Policy," *Journal of Behavioral Economics and Organization* 68 (2008), 632–644. Reprinted with permission from Elsevier Publishers, Amsterdam.

human behavior are based on "armchair theorizing" and not empirical evidence about actual human behavior, criticisms can easily be ignored. But in recent decades, behavioral research has fundamentally changed the field of economics by putting it on a firm experimental basis. In its early days, behavioral economics concentrated on revealing various shortcomings of the standard model of economic choice. Recently, the field has moved from merely reacting against the rational actor model to identifying behavioral regularities that might form the basis for a more realistic model of human decision making. Experiments such as the **Ultimatum Game** and the Public Goods Game have established a number of regularities in human behavior, such as **loss aversion, habituation,** pure altruism, **altruistic punishment,** and inconsistent discounting of the future. These behavioral patterns have been confirmed by neurological experiments showing how behavior is reflected in brain activity.

BEHAVIORAL ECONOMICS AND GAME THEORY

For many years game theory was one of the bastions of orthodoxy in economics. The classic textbook example of the inevitability of selfish behavior is the **Prisoner's Dilemma** (PD). The setting for the PD game is this. The police have captured two people—the Gecko brothers, Seth and Quentin—suspected of committing a serious crime. The case against them is not strong, so a confession is needed from at least one of them. The police put the two brothers in separate rooms and offer them the deal shown in Figure 7.1. If neither confesses they get three years of prison time each. If they both confess they get four years each. If one confesses and the other does not, the confessor gets one year and the non-confessor gets six years. The way the game is framed, it is "rational" for Seth and Quentin to confess no matter what the other one does. Suppose Seth confesses; then Quentin should confess in order to get four years instead of six. Suppose Seth does not confess; then Quentin should also confess in this case to get one year instead of three.

The same logic applies to Seth, who should also confess no matter what Quentin does. This is called a **Nash Equilibrium** (named for Nobel laureate John Nash), which occurs when each player's strategy is optimal, given the

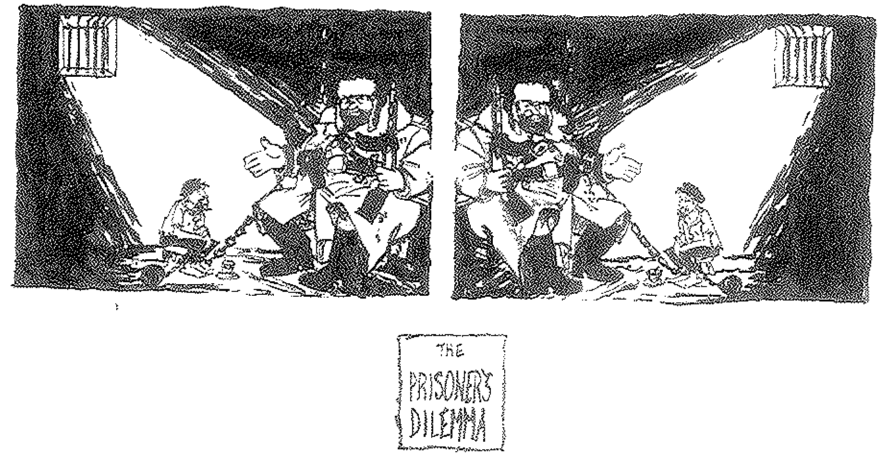
Quentin \ Seth	Confess	Do not confess
Confess	4 years each	1 year for Quentin 6 years for Seth
Do not confess	6 years for Quentin 1 year for Seth	3 years each

Figure 7.1. The Prisoner's Dilemma

strategies of the other players. A player has a **dominant strategy** if that player's best strategy does not depend on what other players do (as in the PD—always confess).

The theoretical result of the PD game, no cooperation, is based on the assumption that there is no interaction between the two players. But in repeated PD games, people tend to cooperate. More surprisingly, even in one-shot anonymous PD experiments, over one-half of the players cooperate (Field 2001). Among the first two people to play the game in the 1950s were the imminent economist and mathematician Armen Alchian and John Williams, a distinguished mathematician at the Rand Corporation. When they cooperated in the one-shot PD game John Nash remarked, "I would have thought them more rational" (quoted in Field 2001).

Another classic game theory experiment is the Ultimatum Game (UG), which has been one of the most important contributions to behavioral economics. Like the Prisoner's Dilemma game before it, the UG helped revolutionize the way economists think about economic decision making. In this game, a leader offers one of two participants a certain sum of money and instructs that participant to share it with the second player. The second



player can either accept the offer or reject it, in which case neither player gets anything. If the players behave according to the model of *Homo economicus*, the first player should offer the minimum amount and the second player should accept any positive offer. Results from the game show, however, that the majority of proposers in Western countries offer between 40 and 50 percent of the total and that offers under 30 percent of the total are usually rejected because they are not "fair" (Nowak, Page, and Sigmund 2000). These results have held up even when the game is played with substantial amounts of real money.

An extensive study using the UG game involved economists and anthropologists who played the game in fifteen different cultures around the world (Henrich et al. 2001, 73). The authors of the study concluded: "The canonical model of self-interested behavior is not supported in *any* society studied." As mentioned above, UG results are consistent in advanced market economies in North American, Europe and Japan. Among other cultures, such as the ones studied by Henrich and his colleagues, the results varied wildly depending on the social norms of the particular cultures studied. Among the whale-hunting Lamalera of Indonesia, 63 percent of the proposers divided the pie equally, and most of those who did not, offered more than 50 percent (the mean offer was 57 percent). In real life, a large catch, always the product of cooperation among many individual whalers, is meticulously divided into predesignated parts and carefully distributed among the members of the community. Among the Au and Gnau of the New Guinea Highlands, many proposers offered more than half the amount they had, and many of these "hyper-fair" offers were rejected! This reflects the Melanesian culture of status-seeking through gift giving. Making a large gift is a bid for social dominance in everyday life in these societies, and rejecting the gift is a rejection of being subordinate.

Another standard game is the Public Goods (PG) game. This game has many variants but a typical version goes something like this: There are ten players and they play the game for ten rounds. On each round each player is given the choice of depositing some amount of money (say 50¢) in "community pool" or keeping a larger amount for himself (say $1). For each player in turn, if the player deposits 50¢ in the common pool, then all the players get 50¢ each. So if all players are cooperative, then each player receives $5 per round (10×50¢) for a total of $50 at the end of the ten-round game. If all players are selfish, they each only get $1 per round or $10 at the end of the game. The catch is that if one player acts selfishly and the other players cooperate, the selfish player gets $5.50 per round (9×50¢+$1) and all the others get $4.50 (9×50¢). So it pays to be a defector (**free rider**) not a cooperator. Standard welfare theory predicts that "rational" players would never cooperate and that each player would take $1 for him- or herself starting with round one of the game. But results of PG games show much more complicated behavior. Typically, the majority of players begin by cooperating but then they change their behavior to defecting when they see others being selfish. If the game is played many times, people build up a sense of trust and there is a return to cooperation. If players are allowed to punish free riders by fining them, the game usually evolves to a cooperative outcome (for a summary of PG games see Gintis 2000, chapter 11, and Fehr and Gächter 2000a).

Results from the Ultimatum Game, the Public Goods game, and other game theoretic experiments show that, in a variety of settings and under a variety of assumptions, other-regarding motives are a better predictor of behavior than those embodied in self-regarding *Homo economicus*. Humans regularly exhibit

a culturally conditioned sense of fairness, and they are willing to enforce cultural norms even at economic cost to themselves. Cross-cultural UG experiments also show that cultural norms vary greatly and that these norms dramatically affect the average amount offered in the game and the rates of rejection (Henrich et al. 2001). Again, a striking result of UG experiments is that the model of rational economic man is not supported in any culture studied (Henrich et al. 2001, 73). Henrich and his colleagues summarize their findings:

> Recent investigations have uncovered large, consistent deviations from the predictions of the textbook representation of *Homo economicus* (Alvin E. Roth et al., 1991; Ernst Fehr and Simon Gächter, 2000[b]; Colin Camerer, 2001). One problem appears to lie in economists' canonical assumption that individuals are entirely self-interested: in addition to their own material payoffs, many experimental subjects appear to care about fairness and reciprocity, are willing to change the distribution of material outcomes at personal cost, and are willing to reward those who act in a cooperative manner while punishing those who do not, even when these actions are costly to the individual. (Henrich et al. 2001, 73)

It took several decades of carefully designed, repeatable experiments to expose rational economic man as an inadequate model of human behavior. Economists and behavioral scientists are now turning their attention to the more difficult but potentially more important task of constructing an alternative to *Homo economicus* based on observed patterns of human behavior.

EMPIRICALLY ESTABLISHED BEHAVIORAL PATTERNS

Experimental results from behavioral economics, evolutionary game theory, and neuroscience have firmly established that human choice is a social, not self-regarding, phenomenon (see the essays in Camerer, Loewenstein, and Rabin 2004). Two broad principles have emerged from the literature: (1) human decision making cannot be accurately predicted without reference to social context, and (2) regular patterns of decision making, including responses to rewards and punishments, can be predicted both within particular cultures and across cultures. Recent research shows that preferences are endogenous, that is, they depend on the individual's personal history, interaction with others, and the social context of individual choice. Several consistent patterns of **endogenous preferences** have been observed.

Loss Aversion

A well-documented behavioral pattern is that people are more concerned about avoiding losses than they are about acquiring gains (Kahneman and Tversky 1991; Knetsch and Sinden 1984). The explicit assumption in economic analysis is that only the absolute magnitude of the change matters, not the direction of the change. Consumers routinely violate this assumption as advertisers know well. Most people would prefer to buy an item listed at $105 with a $5 discount than an item listed at $95 with a $5 surcharge, even though the price is the same in both cases ($100). Loss aversion is closely connected to the **endowment effect.**

The Endowment Effect

The hypothesis that losses are systematically valued more than equivalent gains has been verified in numerous experiments. It seems to be a psychological law that people prefer to keep something they already have compared with acquiring something they do not have (Kahneman and Tversky 1979). Tests of the endowment effect have shown that it is not due to wealth effects, income disparities, strategic behavior, or transactions costs (Kahneman, Knetsch, and Thaler 1991). Dozens of experiments show that preferences depend on the direction of the change, that is, whether people are paid to give up something they have or have to pay to get something they do not have (Knetsch 2007). The psychological model makes good predictions of economic behavior; the rational actor model does not.

Process-Regarding Preferences

People care about process as well as outcome. In designing economic policies, the process of arriving at a decision may be as important for public acceptance as the actual outcome itself. For example, results from the Ultimatum Game, (mean offers and rejection rates) vary significantly according to the process through which money is obtained and the way offers are made. Offers

are substantially lower if proposers win their position by doing well on a quiz (Hoffman et al. 1994). Rejection rates are much lower if respondents are told that the offers were generated by a computer. In the Prisoner's Dilemma game, defection rates are significantly higher if the game is referred to as the "Wall Street Game" rather than the "Community Game." Results from these and numerous other studies in game theory, experimental economics, and behavioral economics show that models that do not take into account social processes such as community norms about fairness may lead to poor predictors of economic behavior.

Time Inconsistency and Hyperbolic Discounting

Time consistency is critical to the standard economic assumption that benefits delivered in the future should be discounted at a fixed rate. But behavioral studies indicate that people discount the near future at a higher rate than the distant future, and they have different discount rates for different kinds of outcomes (Frederick, Loewenstein, and O'Donoghue 2004). Anticipation has been found to be a positive thing in itself and may result in something in the future actually having a *higher* value. This finding is relevant to environmental policies such as preserving national parks and other wildlife areas because individuals may enjoy them more in the future (after retirement, for example) and the anticipation of this is important.

Biased Cultural Transmission

Biased cultural transmission is a theory of innovation diffusion based on the observation that people imitate others whose actions they trust or respect. People use heuristics, mental shortcuts, and rules of thumb to make otherwise complicated decisions. Biased cultural transmission may lead to the widespread adoption of economically inefficient ways of doing things. By *selectively* imitating respected individuals, people may ensure that innovations become established in a community whether or not the innovation is superior to others as determined by cost-benefit calculations (Henrich 2003). The important factor in adoption is the innovation's conformance with established cultural patterns. This has far-reaching implications for the design of economic policies.

Results from game theory and behavioral economics show that preferences are other regarding. People act to affect the well-being of others, positively or negatively, even at significant cost to themselves (Fehr and Gächter 2000a). A sense of fairness, including pure altruism, is a critical factor in economic decisions. This is illustrated in various game theory experiments such as the Public Goods game in which participants are willing to impose, at great cost to themselves, punishments on non-contributors, even in the last round of the game (Bowles and Gintis 2002). These kinds of behavior patterns have important consequences for judgments about human well-being and economic policy design.

THE EVOLUTIONARY BASIS OF HUMAN BEHAVIOR

Also relevant to the study of human decision making is a growing body of evidence from (nonhuman) animal experiments. These experiments show two important things. First, some social animals, such as primates, also have a sense of fairness and a tendency to cooperate. Second, some "lower" animals may behave closer to the rational actor model than humans do. They are self-regarding in evaluating payoffs, they are not susceptible to the **sunk-cost** effect, and they apparently evaluate payoffs according to expected utility theory. As discussed below, animal experiments show that human behavior has an evolutionary basis. They also illuminate the uniqueness of human behavior as occurring within a complex, socially constructed system. Ironically, it is not "rationality" that makes us human, but rather it is the "anomalies" uncovered by behavioral science.

The Behavior of Social Animals

Melis, Hare, and Tomasello (2006) played a cooperation game with chimpanzees at the Ngamba Island Chimpanzee Sanctuary in Uganda. A feeding platform with two metal rings was placed outside a testing room cage with a rope threaded through the rings and the two ends of the rope in the test room cage. If the chimpanzee(s) pulled only on one end of the rope, the rope passed through the rings and the food was not obtained. Only if two chimpanzees pulled together could the platform be pulled forward and the food obtained.

During repeated tests, the chimpanzees were allowed to recruit partners of their own choice, and they quickly learned to recruit those who were the best collaborators. Kin selection is not involved because the chimpanzees at the sanctuary are unrelated orphans from the wild. The authors observe: "Therefore, recognizing when collaboration is necessary and determining who is the best collaborative partner are skills shared by both chimpanzees and humans, so such skills may have been present in their common ancestor before humans evolved their own complex forms of collaboration" (Melis, Hare, and Tomasello 2006, 1297).

As we have seen, there is no place for interactive behavior, including altruism, in the basic Walrasian model. Economists also tend to be skeptical of altruistic behavior because of the free-rider problem. Free riders can out-compete altruists by taking advantage of their generosity. In the standard welfare model they will always out-compete altruists. As an answer to this objection, Henrich and his colleagues (2006, 1767) propose that altruism arose in humans hand in hand with punishing noncooperators. Altruistic punishment—punishing others for violating social norms even at cost to oneself—is one way humans deal with free riders and make cooperation work. Apparently, punishing those who do not cooperate actually stimulates the same pleasure centers in the brain that are activated by, for example, eating something sweet (Vogel 2004, 1131). Some evidence indicates that punishing behavior is present in chimpanzees. In one experiment, semi-wild chimpanzees were fed at a regular time only after all the chimpanzees in the compound came to the feeding station. Latecomers held up the feeding for all the chimps, and these stragglers were punished with hitting and biting.

A sense of what is fair and what is unfair is also present in our primate cousins. Brosnan and de Waal (2003) found that brown capuchin monkeys (*Cebus paella*) exhibit a strong aversion to inequity. In one experiment, monkeys rejected rewards for performing a simple task if they witnessed another monkey receiving a more desirable reward for performing the same task. Pairs of monkeys were trained to exchange a small rock with a human experimenter in return for a piece of cucumber. When one monkey saw the other receiving a more desirable reward (a grape), the first monkey would not only refuse to participate in further exchanges but would frequently refuse to eat the cucumber reward, sometimes even throwing it toward the human experimenter. Brosnon and de Wall (2003, 299) write: "People judge fairness based both on the distribution of gains and on the possible alternatives to a given outcome. Capuchin monkeys, too, seem to measure reward in relative terms, comparing their own rewards with those available, and their own efforts with those of others."

In terms of standard economic theory, the question is not whether humans (or other animals) are selfish or altruistic but whether they are *other regarding*. As shown in Chapter 1, if individuals evaluate their payoffs based on what others get, this violates the conditions for Pareto efficiency in the standard model. Other-regarding behavior may be altruistic, envious, or any other socially conditioned response to others. For example, researchers found that in cooperation games with an opportunity to punish, subjects from Belarus and Russia punished not only defectors but also strong cooperators (Vogel 2004)!

Are "Lower" Animals More Rational than Humans?

The view of the human rational actor as a highly evolved decision maker has also taken a blow from studies of animals with more limited cognitive ability. Regarding the claims for human rationality, it is ironic that a large body of evidence suggests that some lower animals act more in accordance with the economic model of rational choice than humans do. In a classic experiment, Harper (1982) tested the ability of a flock of ducks to achieve a stable Nash Equilibrium when fed balls of bread. Every morning two researchers would stand on the bank of the pond where the ducks were and throw out five-gram dough balls at different intervals. Expected utility theory would predict that the ducks would distribute themselves between the two feeders in such a way that $N1 / r1 = N2 / r2$, where Ni is the number of ducks and ri is the expected (bread) payoff from standing in front of one of the feeders. So if there are thirty-three ducks participating and if one experimenter throws a five-gram ball of dough every fifteen seconds and the other experimenter throws a five-gram ball of dough every thirty seconds, there should be twenty-two ducks in front of the first experimenter and eleven in front of the other. And in fact this is what happened. The ducks rearranged their numbers efficiently as the payoffs were changed. Furthermore, when the experimenters changed the speed of throwing the dough balls, the ducks would quickly and efficiently readjust their numbers. Glimcher (2002, 329) writes:

> One thing that was particularly striking about this result was the speed at which the ducks achieved this assortment. After 90 seconds of breadball throwing, as few as ten breadballs have been dispersed. Long before half the ducks have obtained even a single breadball, they have produced a precise equilibrium solution.

A well-known violation of rationality is the sunk-cost effect. Ignoring unrecoverable past expenditures is one the common admonishments for students learning to "think like an economist," that is to behave in a sophisticated rational way (Frank and Bernanke 2004). But once again, actual human behavior consistently deviates from the rational actor ideal. A number of experiments have demonstrated that human decisions are strongly influenced by sunk costs. It appears, however, that ignoring sunk costs is a characteristic of the behavior of lower animals but not of humans (Arkes and Ayton 1999). Fantino (2004) performed a simple investment experiment with college students and pigeons. Both were rewarded with money or food for pressing a computer keyboard an undetermined number of times until an award was given. Pressing some of the keys resulted in an award whereas pressing others produced no reward. The experiment was designed to model a bad investment in which the chances of success diminished as the number of responses increased. The more times a key was pressed with no reward forthcoming, the less likely further pressing would produce an award. In the experiment, the pigeons quickly switched from one key to another if an award failed to appear, whereas the students kept repeatedly pressing the same key—indicating that pigeons were less susceptible than students to the sunk-cost effect. In another sunk-cost experiment, Maestripieri and Alleva (1991) tested the behavior of mother mice in defending their young, and they found that the aggressiveness of their defensive behavior depended on the number of offspring in the litter rather than the amount of time invested in caring for them.

The animal behavior literature, together with observations of human behavior, suggests that letting sunk costs influence decision making is a trait that must have something to do with uniquely human characteristics such as the presence of complex capital investments and complex institutions in human societies. It is sometimes argued that although individuals may exhibit irrational behavior, such behavior can be corrected in groups (as in the rational expectations literature). In fact, research shows that groups are probably more susceptible to the sunk-cost effect than are individuals (Whyte 1993).

NEUROSCIENCE CONFIRMATION OF BEHAVIORAL REGULARITIES

A rapidly growing field is **neuroeconomics**, that is, identifying regularities in brain activity corresponding to specific human economic decisions. These neurological findings may not add anything new to the catalog of behavioral patterns observed by behavioral economics, but they do show that they are more than anomalies. These observed behaviors are not random mistakes but rather are a part of the neurological organization of the human brain.

Habituation

It has long been known that two groups of neurons, in the *ventral tegmental* and the *substantia nigra pars compacta* areas, and the dopamine they release are critical for reinforcing certain kinds of behavior (Glimcher, Dorris, and Bayer 2005; Schultz 2002). Schultz (2002) measured the activity of these neurons while thirsty monkeys sat quietly and listened for a tone that was followed by a squirt of fruit juice into their mouths. After a period of a fixed, steady amount of juice, the amount of juice was doubled without warning. The rate of neuron firing went from about three per second to eighty per second. As this new magnitude of reward was repeated, the firing rate returned to the baseline rate of three firings per second. The opposite happened when the reward was reduced without warning. The firing rate dropped dramatically but then returned to the baseline rate of three firings per second.

The Framing Effect

Consistency in choice is the hallmark of rational economic man, and it implies that the evaluation of choices is unaffected by the manner in which the choices are framed. This view was challenged by Kahneman and Tversky (1979) in their formulation of "prospect theory," that is, people evaluate changes in terms of a reference point. The "framing effect" means that the frame of reference may change according to how a particular choice is presented, and this will affect the payoff decision. This effect has been confirmed in numerous experiments, and it too seems to have a neurological basis. De Martino and his colleagues (2006) used functional magnetic resonance imaging (fMRI) to look at the neurological effects of framing in a simple experiment. A group of twenty British subjects were asked to choose between identical outcomes framed differently. They were told that they would initially receive £50. They then had to choose between a "sure" option and a "gamble" option. The sure option was presented in two ways, either as a gain (say, keep £20 of the £50) or as a loss (say, lose £30 of the £50). The gamble option was presented in the same way in both cases—a pie chart showing the probability of winning or losing. People responded differently depending on how the question was framed, and this was reflected in fMRI images. Different parts of the brain lit up depending on how the question was framed.

The fact that the framing effect found in this experiment had a neurological basis was confirmed: "Our data provide a neurobiological account of the framing effect, both within and across individuals. Increased activation in the amygdale was associated with subjects' tendency to be risk-averse in the Gain frame and risk-seeking in the Loss frame, supporting the hypothesis that the framing effect is driven by an affect heuristic underwritten by an emotional system" (De Martino et al. 2006, 686).

The neural basis for loss aversion was also confirmed by Tom and colleagues (2007). They found that in order for people to accept a fifty-fifty gamble the potential gain needs to be twice as high as the potential loss. They discovered that the brain regions that evaluated potential gains and losses were more sensitive to losses. Also, between-subject differences in loss aversion reflected between-subject differences in neural responses.

Threshold Effects

In a study of how rhesus monkeys respond in a color matching experiment, Schall and Thompson (1999) found a correlation between neural firing rates and making a physical movement. Thirsty monkeys were trained to stare at a cross in the center of a blank display. Then a circle of eight spots were illuminated, seven in one color and the eighth in another. If the monkey moved his gaze to look at the "oddball" color he was rewarded with a squirt of juice. When the oddball color was identified, neural firing rates began to increase at the location in the brain encoding the oddball. Only after the neural firing rate passes an apparently fixed threshold did the monkey move his gaze. Glimcher, Dorris, and Bayer (2005) postulate that the decision-making brain forms a kind of topological map that encodes something such as the relative expected gains of each possible choice. Actually making a choice (taking an action) depends on the strength of the signal relating to that particular action (the neural firing threshold).

The finding that some animals behave according to the rational actor model can be interpreted in very different ways. The view is widespread that animal behavior justifies the economic rational actor model. Gintis (2007, 7), for example, argues that the assumption of choice consistency among humans is justified by animal behavior: "Economic and biological theory thus have a natural affinity; the choice consistency on which the rational actor model of economic theory depends is rendered plausible by biological evolutionary theory, and the optimization techniques pioneered by economic theorists are routinely applied and extended by biologists in modeling the behavior of a vast array of organisms." Others take the view that animal studies show that the rational choice model is inappropriate to describe all but the simplest kinds of human decision making. Camerer, Loewenstein, and Prelec (2005, 55) write:

> Our view is that establishing a neural basis for some rational choice principles will not necessarily vindicate the approach as widely applied to *humans*. . . . Ironically, rational choice models might therefore be most useful in thinking about the simplest kinds of decisions humans and other species make—involving perceptual tradeoffs, motor movements, foraging for food and so forth—and prove least useful in thinking about abstract, complex, long-term tradeoffs which are the traditional province of economic theory.

Far from describing higher-order, complex human behavior, the axiomatic rational choice model strips away everything that makes humans unique as highly intelligent social animals. Nelson (2005, 264) puts it succinctly:

> Defining economics as the study of *rational choice*, neoclassical economics treats human physical bodies, their needs, and their evolved actual psychology of thought and action as rather irrelevant. The notion that humans are created as rational decision-makers is, from a physical anthropology point of view, just as ludicrous as the notion that humans were created on the sixth day.

Our very complex, other-regarding, altruistic, empathetic behavior is what makes humans unique, and understanding this behavior is the key to formulating effective economic policies having complicated and long-lasting consequences.

WHY DOES ALL THIS MATTER? THE IMPORTANCE OF BEHAVIORAL ASSUMPTIONS

At the core of neoclassical welfare theory is the rational actor model of human behavior. Individuals act to maximize utility according to consistent, constant, well-ordered, and well-behaved preferences. In the rational actor model, preferences are **exogenous**, that is, other individuals or social institutions do not influence them. The argument for using individual preferences as the starting point is a powerful one. It is a good thing for individuals to have what they want, and each individual is the best judge of what he or she wants. According to Randall (1988, 217), economists are "doggedly nonjudgmental about people's preferences." But are they? In fact, by forcing individual preferences through the narrow funnel of rational choice theory, economists are denying individuals a whole range of choices falling under the rubric of endogenous preferences.

For example, in surveys of consumer preferences for environmental services, information is collected that routinely violates the axioms of consumer choice theory. People express ethical concerns based on group norms (Fehr and Gächter 2000a; Gowdy and Seidel 2004), and considerable evidence exists that people value the medium and distant futures about the same (hyperbolic discounting) (Laibson 1997). But collected information about consumer

Expressed preferences		Eliminated by assuming		Policy implications
Lexicographic preferences	→	Continuity	→	Everything is tradable
Inconsistent discounting	→	Time consistency	→	Straight-line discounting
Endowment effect	→	Symmetric rationality	→	WTA = WTP
Other-regarding preferences	→	Independent choices	→	Relative income does not matter
Process-regarding preferences	→	Outcome-regarding preferences	→	Process does not matter

Figure 7.2. Preference filtering in the Walrasian model

attitudes is filtered by economists through the axioms of consumer choice to fit the stylized "facts" of neoclassical welfare economics (as shown in Figure 7.2). Thus subjectivism and values enter economics in a non-explicit way that is more dangerous than making explicit value judgments.

These filters take a variety of forms. For example, in surveys using the **contingent valuation method** (CVM), "protest bids" are very common. These may be in the form of extreme bids of zero or infinity. One reason for these bids is the existence of **lexicographic preferences**, that is, people may place absolute values on environmental preservation and refuse to make trade-offs between environmental features and money. Spash and Hanley (1995) found that approximately 25 percent of respondents in CVM surveys refused to consider the concept of trading income changes with changes in environmental quality. In many CVM surveys, these bids are excluded from the analysis thereby disenfranchising those respondents. A common finding is the existence of lexicographic preferences. In this case people have an absolute preference for something and are unwilling to except any substitute for that particular thing. A recent trend in CVM studies is to filter out lexicographic preferences by designing surveys to elicit market-compatible responses. Bid cards, for example, restrict choices in CVM surveys to a given set of offers, thus forcing them to conform to the normative assumptions of the investigator. Willingness to pay measures of the value of environmental goods are routinely used

even though a considerable amount of evidence shows the correct measure to use is the willingness to avoid losses.

The theoretical and empirical breakthroughs described in this chapter and in Chapter 6 are beginning to significantly influence the theory and methods of economics. However, although economists rightly point out that mainstream theorists have extended the neoclassical paradigm far beyond the limits of traditional welfare analysis, the leading textbooks, and the policy recommendations of most economists, continue to rely almost exclusively on the basic framework of the Walrasian system. The prospects for a reorientation of economic theory and policy are discussed in the next two chapters.

GLOSSARY

Altruistic punishment—The willingness of economic actors to punish others who violate perceived social norms, even at substantial cost to themselves.

Biased cultural transmission—A theory of innovation diffusion based on the observation that people imitate others whose actions they trust or respect.

Contingent valuation method (CVM)—Soliciting consumer preferences using sophisticated questionnaires and interviewing techniques. The design and interpretation of CVM surveys are based on the standard economic model of rationality and consistency in choices.

Dominant strategy—A player's best strategy that does not depend on what the other players do.

Endogenous preferences—Preferences that depend upon (are endogenous to) social norms.

Endowment effect—People prefer to keep something they already have compared with acquiring something they do not have.

Exogenous preferences—Preferences that are independent of (exogenous to) social norms.

Framing effect—Changing the frame of reference may change the evaluation of choices.

Free rider—A selfish individual who takes advantage of the generosity of others. Given the behavioral assumptions of the standard model, free riders will always drive out altruists in a competitive situation.

Habituation—The perceived positive or negative benefits of gains or losses, if repeated, tend to disappear over time.

Hyperbolic discounting—People tend to discount the future more heavily in the immediate future than they do in the distant future.

Lexicographic preferences—These occur when a consumer infinitely prefers one good to another. In this case a commodity bundle containing more of the lexicographic good will be preferred to any other commodity bundle. Substitutability does not exist.

Loss aversion—People are willing to pay more to avoid the loss of something than they are willing to pay to gain something they do not have.

Nash Equilibrium—Occurs when each player's strategy is optimal, given the strategy of the other players.

Neuroeconomics—Testing the effects of economic decision making on brain activity using fMRI imaging or other measures of neurological activity.

Prisoner's Dilemma—The classic game theory experiment purportedly showing the inevitability of noncooperative behavior when players cannot interact.

Process-regarding preferences—People care about the process through which an outcome is obtained as well as the outcome itself.

Sunk-cost effect—Letting unrecoverable past expenditures influence the decision-making process. This is "irrational" in standard theory.

Threshold effect—In neuroscience experiments, actually making a choice (taking an action) depends on the strength of the signal relating to that particular action (the neural firing threshold). No action is taken until a certain threshold is reached.

Time inconsistency—People do not discount the future in a consistent way. People use different discount rates for different ranges of time.

Ultimatum Game—One of the first modern game theory experiments showing the pervasiveness of other-regarding behavior (altruistic punishment).

REFERENCES AND FURTHER READING

Altruism and Altruistic Punishment

Field, A. 2001. *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity.* Ann Arbor: University of Michigan Press.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. 2006. Costly punishment across human societies. *Science* 312, 1767–1770.

Vogel, G. 2004. The evolution of the golden rule. *Science* 303, 1128–1131.

Animal Behavior and Evolutionary Psychology

Arkes, H., and P. Ayton. 1999. The sunk cost and concorde effects: Are humans less rational than lower animals? *Psychological Bulletin* 125, 591–600.

Brosnan, S., and F. de Wall. 2003. Monkeys reject unequal pay. *Nature* 425, 297–299.

Harper, D. 1982. Competitive foraging in mallards: "Ideal free" ducks. *Animal Behavior* 30, 575–584.

Maestripieri, D., and E. Alleva. 1991. Litter defense and parental investment allocation in house mice. *Behavioral Processes* 56, 223–230.

Melis, A., B. Hare, and M. Tomasello. 2006. Chimpanzees recruit the best collaborators. *Science* 311, 1297–1300.

Behavioral Economics, Theory, and Policy

Bowles, S., and H. Gintis. 2002. Social capital and community governance. *Economic Journal* 112, F419–F436.

Camerer, C., G. Loewenstein, and M. Rabin (eds.). 2004. *Advances in Behavioral Economics.* Princeton, NJ: Princeton University Press.

Fantino, E. 2004. Behavior-analytic approaches to decision making. *Behavioral Processes* 66, 279–288.

Gowdy, J., and I. Seidl. 2004. Economic man and selfish genes: The relevance of group selection to economic policy. *Journal of Socio-Economics* 33, 343–358.

Henrich, J. 2003. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization* 53, 3–35.

Kahneman, D. 2003. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review* 93(5), 1449–1475.

Knetsch, J. 2007. Biased valuations, damage assessments, and policy choices: The choice of measure matters. *Ecological Economics* 63, 684–689.

Discounting and Time Preference

Frederick, S., G. Loewenstein, and T. O'Donoghue. 2004. Time discounting and time preference: A critical review. In C. Camerer, G. Lowenstein, and M. Rabin (eds.), *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press, 162–222.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112, 443–477.

Games

Fehr, E., and S. Gächter. 2000a. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980–994.

Fehr, E., and S. Gächter. 2000b. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14(3), 159–181.

Gintis, H. 2000. *Game Theory Evolving*. Princeton, NJ: Princeton University Press.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2001. In search of Homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review* 91, 73–78.

Hoffman, E., K. McCabe, K. Shachat, and V. Smith. 1994. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7, 346–380.

Nowak, M., K. Page, and K. Sigmund. 2000. Fairness versus reason in the ultimatum game. *Science* 289, 1773–1775.

Roth, A., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study. *American Economic Review* 81(5), 1068–1095.

Neuroeconomics and Neuroscience

Camerer, C., G. Loewenstein, and D. Prelec. 2005. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43, 9–64.

De Martino, B., D. Kumaran, B. Seymour, and R. Dolan. 2006. Frames, biases, and rational decision-making in the human brain. *Science* 313, 684–687.

Glimcher, P. 2002. Decisions, decisions, decisions: Choosing a biological science of choice. *Neuron* 36, 323–332.

Glimcher, P., M. Dorris, and H. Bayer. 2005. Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior* 52, 213–256.

Schall, J., and K. Thompson. 1999. Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience* 22, 241–259.

Schultz, W. 2002. Getting formal with dopamine and reward. *Neuron* 36, 241–263.

Tom, S., C. Fox, C. Trepel, and R. Poldrack. 2007. The neural basis of loss aversion in decision-making under risk. *Science* 315, 515–518.

Other References Cited

Camerer, C. 2001. *Behavioral Economics*. Princeton, NJ: Princeton University Press.

Frank, R., and B. Bernanke. 2004. *Principles of Economics*, 2nd ed. New York: McGraw-Hill.

Gintis, H. 2007. A framework for the integration of the behavioral sciences. *Behavioral and Brain Sciences* 30, 1–16.

Kahneman, D., J. Knetsch, and R. Thaler. 1991. The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 5, 193–206.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.

Kahneman, D., and A. Tversky. 1991. Loss aversion in riskless choices: A reference dependent model. *Quarterly Journal of Economics* 106, 1039–1061.

Knetsch, J., and J. Sinden. 1984. Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *Quarterly Journal of Economics* 99, 507–521.

Nelson, J. 2005. Is economics a natural science? *Cosmos and History: The Journal of Natural and Social Philosophy* 2, 261–269.

Randall, A. 1988. What mainstream economists have to say about the value of biodiversity. In E. O. Wilson (ed.), *Biodiversity*. Washington, DC: National Academy Press, 217–223.

Spash, C., and N. Hanley. 1995. Preferences, information, and biodiversity preservation. *Ecological Economics* 12, 191–208.

Veblen, T. 1898. Why is economics not an evolutionary science? *Quarterly Journal of Economics* 12, 373–397.

Whyte, G. 1993. Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes* 54, 430–455.

8

COST-BENEFIT ANALYSIS OLD AND NEW

Benefit-cost practice and policy analyses were greatly improved when non-market, or non-pecuniary, values were included—a change of relatively recent vintage, and still not universal. The evidence now suggests further improvements could be made by taking account of more realistic behavioral assumptions in valuing gains and losses.

—*Jack Knetsch, "Gains, Losses, and the US-EPA* Economic Analyses Guidelines: *A Hazardous Product,"* Environmental & Resource Economics *32 (2005), 110*

THE WELFARE FOUNDATIONS OF COST-BENEFIT ANALYSIS

The economic theory behind standard cost-benefit analysis (CBA) begins with the basic assumptions about human preferences discussed in Part One of this book. Economic value arises solely from human preferences and these preferences are stable over time (Stigler and Becker 1977). It is the task of econometricians to uncover these hidden but stable preferences through surveys (such as the **contingent valuation** method) or by imputing prices for goods not directly traded in markets (**hedonic pricing**). The object is to identify potential Pareto improvements so that public policies can be designed to correct these inefficiencies (market failures). Whether or not market choices, or pseudo-market choices, identified by CBA accurately reflect what is "best" for society depends critically on the validity of the assumptions of the rational actor model.

This chapter was originally published in a somewhat different form in John M. Gowdy, "Toward an Experimental Foundation for Benefit-Cost Analysis," *Ecological Economics* 63 (2007), 649–655. Reprinted with permission from Elsevier Publishers, Amsterdam.