4 Valuing the Environment: Methods

For it so falls out; That what we have we prize not to the worth;
Whiles we enjoy it, but being lack'd and lost; Why, then we rack
the value; then we find; The virtue that possession would not show us;
Whiles it was ours.

—William Shakespeare, *Much Ado About Nothing*

Introduction

Soon after the *Exxon Valdez* oil tanker ran aground on the Bligh Reef in Prince William Sound off the coast of Alaska on March 24, 1989, spilling approximately 11 million gallons of crude oil, the Exxon Corporation (now Exxon Mobil) accepted the liability for the damage caused by the leaking oil. This liability consisted of two parts: (1) the cost of cleaning up the spilled oil and restoring the site insofar as possible, and (2) compensation for the damage caused to the local ecology. Approximately $2.1 billion was spent in cleanup efforts and Exxon also spent approximately $303 million to compensate fishermen whose livelihoods were greatly damaged for the five years following the spill.[1] Litigation on environmental damages settled with Exxon agreeing to pay $900 million over 10 years. The punitive damages phase of this case began in May 1994. In January 2004, after many rounds of appeals, the U.S. District Court for the State of Alaska awarded punitive damages to the plaintiffs in the amount of $4.5 billion.[2] This amount was later cut almost in half to $2.5 billion and in 2008 the Supreme Court ruled that even those punitive damages were excessive based on maritime law and further argued that the punitive damages should not exceed the $507 million in compensatory damages already paid.[3]

In the spring of 2010, the Deepwater Horizon, a BP well in the Gulf of Mexico, exploded and began spewing an Exxon Valdez–size oil spill every 4–5 days. By the time the leaking well was capped in August 2010, more than 200 million gallons had been spread through the Gulf of Mexico, almost 20 times greater than the Exxon Valdez spill.

[1]U.S. District Court for the State of Alaska, Case Number A89-0095CV, January 28, 2004.

[2]Ibid.

[3]*Exxon Shipping Company v. Baker.*

What are the economic damages from spills like these that threaten the loss of valuable fisheries and tourism, as well as many other individual biological species, including several endangered turtle species and hundreds of bird species? Thousands of birds have been found dead in the Gulf since the BP spill, for example.[4] Interestingly, the *Exxon Valdez* spill triggered pioneering work focused on providing monetary estimates of environmental damages, setting the stage for what is today considered standard practice for nonmarket valuation.

In Chapter 3 we examined the basic concepts used by economists to calculate this damage. Yet implementing these concepts is far from a trivial exercise. While the costs of cleanup are fairly transparent, estimating the damage is more complex. For example, how was the number $900 million in damages in the Exxon case arrived at?

In this chapter we explore how we can move from the general concepts to the actual estimates of compensation required by the courts. A series of special techniques has been developed to value the benefits from environmental improvement or, conversely, to value the damage done by environmental degradation. Special techniques were necessary because most of the normal valuation techniques that have been used over the years cannot be applied to environmental resources. Benefit–cost analysis requires the monetization of all relevant benefits and costs of a proposed policy or project, not merely those where the values can be derived from market transactions As such, it is also important to monetize those environmental goods and services that are not traded in any market. Even more difficult to grapple with are those nonmarket benefits associated with passive-use or nonuse value, topics explored below.

Why Value the Environment?

While it may prove difficult, if not impossible, to place an accurate value on certain environmental amenities, not doing so leaves us valuing them at $0. Will valuing them at $0 lead us to the best policy decisions? Probably not, but that does not prevent controversy from arising over attempts to replace $0 with a more appropriate value (Debate 4.1).

Many federal agencies require benefit–cost analysis for decision making. Ideally, the goal is to choose the most economically desirable projects, given limited budgets. A 1982 amendment to the Endangered Species Act, for example, required benefit–cost analysis for the listing of a species. This requirement was subsequently relaxed, however, due to a lack of defensible benefits measurements. Estimation of benefits and costs is also used for natural resources damage assessments, such as for oil spills. The Federal Energy and Regulatory Commission (FERC) requires benefit–cost analysis for dam relicensing applications. These analyses, however, frequently fail to incorporate important nonmarket values associated with rivers. If the analysis does not include all the appropriate values, the results will be flawed. Have we made progress?

[4]http://www.fws.gov/home/dhoilspill/pdfs/Bird%20Data%20Species%20Spreadsheet%2012142010.pdf

DEBATE 4.1

Should Humans Place an Economic Value on the Environment?

Arne Naess, the late Norwegian philosopher, used the term *deep ecology* to refer to the view that the nonhuman environment has "intrinsic" value, a value that is independent of human interests. Intrinsic value is contrasted with "instrumental" value in which the value of the environment is derived from its usefulness in satisfying human wants.

Two issues are raised by the Naess critique: (1) What is the basis for the valuing of the environment? and (2) how is the valuation accomplished? The belief that the environment may have a value that goes beyond its direct usefulness to humans is in fact quite consistent with modern economic valuation techniques. As we shall see in this chapter, economic valuation techniques now include the ability to quantify a wide range of "nonuse" values as well as the more traditional "use" values.

Controversies over how the values are derived are less easily resolved. As described in this chapter, economic valuation is based firmly upon human preferences. Proponents of deep ecology, on the other hand, would argue that allowing humans to determine the value of other species would have no more moral basis than allowing other species to determine the value of humans. Rather, deep ecologists argue, humans should only use environmental resources when necessary for survival; otherwise, nature should be left alone. And, because economic valuation is not helpful in determining survival necessity, deep ecologists argue that it contributes little to environmental management.

Those who oppose all economic valuation face a dilemma: when humans fail to value the environment, it may be assigned a default value of zero in calculations designed to guide policy. A value of zero, however derived, will tend to justify a great deal of environmental degradation that could not be justified with proper economic valuation. As a 1998 issue of *Ecological Economics* demonstrated, a number of environmental professionals now support economic valuation as a way to demonstrate the enormous value of the environment to modern society. At the very least, support seems to be growing for the proposition that economic valuation can be a very useful means of demonstrating when environmental degradation is senseless, even when judged from a limited anthropomorphic perspective.

Sources: R. Costanza et al., "The Value of Ecosystem Services: Putting the Issues in Perspective." ECOLOGICAL ECONOMICS, Vol. 25, No. 1 (1998), pp. 67–72; and Gretchen Daily and Katherine Ellison, THE NEW ECONOMY OF NATURE: THE QUEST TO MAKE CONSERVATION PROFITABLE (Washington, DC: Island Press, 2003).

Valuing Environmental Services: Pollination as an Example

Pollination is one example of a valuable ecosystem service with multiple benefits, including nonmarket impacts such as aiding in genetic diversity, ecosystem resilience and nutrient cycling, as well as direct economic impacts of increasing the productivity of agricultural crops. Many agricultural crops rely on bee pollination.

Consider one domestic example. Some 1,000,000 honeybee hives, or more than 40 percent of all the beehives in the United States, are required for cross-pollination of the $2 billion almond crop in California. When the almond trees flower, managed honeybee hives are moved by flatbed trucks to the San Joaquin Valley to provide sufficient bees to pollinate the crop (Ratnieks and Carreck, 2010).

Unfortunately this important ecosystem service may be in jeopardy. In 2006, the popular press began reporting on what has been called Colony Collapse Disorder, an unexplained disappearance of honeybee colonies. Beekeeper surveys suggest that 33 percent of honeybee colonies in the United States died in the winter of 2010. While the exact causes are, as of yet, unknown, multiple causes are likely to blame.

What would be the global cost of losing or reducing this valuable ecosystem service? One study argues that possible future shortages are likely to have quite different economic impacts around the globe (Example 4.1).

EXAMPLE 4.1

Valuing Ecosystem Services: Pollination, Food Security, and the Collapse of Honeybee Colonies

Utilizing a multi-region, computable general equilibrium (CGE) model of agricultural production and trade, Bauer and Wing (2010) examined the global economic impacts of pollinator declines. CGE models produce numerical assessments of economy-wide consequences of various events or programs. This general equilibrium model includes both direct effects on the crop sector and the indirect, noncrop effects. The value of a CGE model over other methods previously utilized in the literature to value pollination services lies in its ability to "track changes in prices across multiple interrelated markets in a consistent fashion . . . " (p. 377). Using this model the authors can estimate not only the impacts, but also how the impacts are affected by the presence of different substitutes for pollination services.

Since fruits, vegetables, and nuts are most dependent on pollination (for some crops pollination is essential), they begin by identifying the pollination dependency of various world crops and how that production could be affected by shortages of pollination services (when the demand for pollinator services exceeds the supply).

They find that the annual, global losses to the crop sector, attributable to a decline in pollination services, are estimated to be $10.5 billion, but economy-wide losses (noncrop sectors) are estimated to be much larger, namely $334 billion.

Examples of the noncrop sectors that are impacted by pollinator declines include livestock since some pollinated plants are used as feed, processed food (e.g., Mrs. Smith's Blueberry Pie, Sara Lee Pecan Rolls), and chemicals such as fertilizers and pesticides.

They also show that some regions of the world, especially western Africa, are likely to suffer disproportionately. This is due not only to the fact that pollinator-dependent crops make up a relatively large share of western Africa's agricultural output, but also to the relative importance of the agriculture sector in the African economy. Whether mechanized or manual pollination could reduce the potential losses remains an open question.

Source: Dana Marie Bauer and Ian Sue Sing, "Economic Consequences of Pollinator Declines: A Synthesis." AGRICULTURAL AND RESOURCE ECONOMICS REVIEW, 39(3): October 2010, pp. 368–383.

Ratnieks and Carreck go on to speculate about potential future losses and ask the important question,

> *Is the future of U.S. commercial beekeeping going to be based on pollinating a few high-value crops? If so, what will be the wider economic cost arising from crops that have modest yield increases from honey bee pollination? These crops cannot pay large pollination fees but have hitherto benefited from an abundance of honey bees providing free pollination.*

Costs to other parts of the world could even be significantly higher than those in the United States.

Valuation

While the valuation techniques we shall cover can be applied to both the damage caused by pollution and the services provided by the environment, each context offers its own unique problems. We begin our investigation of valuation techniques by exposing some of the difficulties associated with one of those contexts, pollution control.

In the United States, damage estimates are not only used in the design of policies, but, as indicated in the opening paragraphs of this chapter, they have also become important in the courts. Some basis for deciding the magnitude of liability awards is necessary.[5]

The damage caused by pollution can take many different forms. The first, and probably most obvious, is the effect on human health. Polluted air and water can cause disease when ingested. Other forms of damage include loss of enjoyment from outdoor activities and damage to vegetation, animals, and materials.

Assessing the magnitude of this damage requires (1) identifying the affected categories; (2) estimating the physical relationship between the pollutant emissions (including natural sources) and the damage caused to the affected categories; (3) estimating responses by the affected parties toward averting or mitigating some portion of the damage; and (4) placing a monetary value on the physical damages. Each step is often difficult to accomplish.

Because the data used to track down causal relationships do not typically come from controlled experiments, identifying the affected categories is a complicated matter. Obviously we cannot run large numbers of people through controlled experiments. If people were subjected to different levels of some pollutant, such as carbon monoxide, so that we could study the short-term and long-term effects, some might become ill and even die. Ethical concern precludes human experimentation of this type.

This leaves us essentially two choices. We can try to infer the impact on humans from controlled laboratory experiments on animals, or we can do statistical analysis of differences in mortality or disease rates for various human populations living in

---

[5]The rules for determining these damages are defined in Department of Interior regulations. See 40 Code of Federal Regulations 300:72–74.

polluted environments to see the extent to which they are correlated with pollution concentrations. Neither approach is completely acceptable.

Animal experiments are expensive, and the extrapolation from effects on animals to effects on humans is tenuous at best. Many of the significant effects do not appear for a long time. To determine these effects in a reasonable period of time, test animals are commonly subjected to large doses for relatively short periods. The researcher then extrapolates from the results of these high-dosage, short-duration experiments to estimate the effects of low-dose, long-duration exposure to pollution on a human population. Because these extrapolations move well beyond the range of experimental experience, many scientists disagree on how the extrapolations should be accomplished.

Statistical studies, on the other hand, deal with human populations subjected to low doses for long periods, but, unfortunately, they have another set of problems—correlation does not imply causation. To illustrate, the fact that death rates are higher in cities with higher pollution levels does not prove that the higher pollution caused the higher death rates. Perhaps those same cities averaged older populations, which would tend to lead to higher death rates. Or perhaps they had more smokers. The existing studies have been sophisticated enough to account for many of these other possible influences but, because of the relative paucity of data, they have not been able to cover them all.

The problems discussed so far arise when identifying whether a particular effect results from pollution. The next step is to estimate how strong the relationship is between the effect and the pollution concentrations. In other words, it is necessary not only to discover *whether* pollution causes an increased incidence of respiratory disease, but also to estimate *how much* reduction in respiratory illness could be expected from a given reduction in pollution.

The nonexperimental nature of the data makes this a difficult task. It is not uncommon for researchers analyzing the same data to come to remarkably different conclusions. Diagnostic problems are compounded when the effects are synergistic—that is, when the effect depends, in a nonadditive way, on what other elements are in the surrounding air or water at the time of the analysis.

Once physical damages have been identified, the next step is to place a monetary value on them. It is not difficult to see how complex an undertaking this is. Consider, for example, the difficulties in assigning a value to extending a human life by several years or to the pain, suffering, and grief borne by both a cancer victim and the victim's family.

How can these difficulties be overcome? What valuation techniques are available not only to value pollution damage, but also to value the large number of services that the environment provides?

Types of Values

Economists have decomposed the total economic value conferred by resources into three main components: (1) use value, (2) option value, and (3) nonuse value. Use value reflects the direct use of the environmental resource. Examples include fish harvested from the sea, timber harvested from the forest, water extracted from a stream for irrigation, even the scenic beauty conferred by a natural vista. If you used one of your senses to experience the resource—sight, sound, touch, taste, or smell—then you have *used* the resource. Some of these uses are called *passive-use values* or *nonconsumptive use values* if the resource is not actually used up (consumed) in the process of experiencing it. Pollution can cause a loss of use value, such as when air pollution increases the vulnerability to illness, an oil spill adversely affects a fishery, or when smog enshrouds a scenic vista.

A second category of value, the option value, reflects the value people place on a future ability to use the environment. Option value reflects the willingness to pay to preserve the option to use the environment in the future even if one is not currently using it. Whereas use value reflects the value derived from current use, option value reflects the desire to preserve the potential for possible future use. Are you planning to go to Yellowstone National Park next summer? Perhaps not, but would you like to preserve the option to go someday?

The third and final category of value, nonuse value, reflects the common observation that people are more than willing to pay for improving or preserving resources that they will never use. One type of nonuse values is a *bequest value*. Bequest value is the willingness to pay to ensure a resource is available for your children and grandchildren. A second type of nonuse value, a pure nonuse value, is called *existence value*. Existence value is measured by the willingness to pay to ensure that a resource continues to exist in the absence of any interest in future use. The term *existence value* was coined by economist John Krutilla in his now-famous quote, "There are many persons who obtain satisfaction from mere knowledge that part of wilderness North America remains even though they would be appalled by the prospect of being exposed to it."[6]

When the Bureau of Reclamation began looking at sites for dams near the Grand Canyon, groups such as the Sierra Club rose up in protest of the potential loss of this unique resource. When Glen Canyon was flooded by Lake Powell, even those who never intended to visit recognized this potential loss. Because this value does not derive either from direct use or potential use, it represents a very different category of value.

These categories of value can be combined to produce the total willingness to pay (TWP):

$$\text{TWP} = \text{Use Value} + \text{Option Value} + \text{Nonuse Value}.$$

Since nonuse values are derived from motivations other than personal use, they are obviously less tangible than use values. Total willingness to pay estimated without nonuse values, however, will be less than the minimum amount that would be required to compensate individuals if they are deprived of this environmental asset. Furthermore, as Example 4.2 makes clear, estimated nonuse values can be quite large. Therefore, it is not surprising that they are controversial. Indeed when

---

[6]Krutilla, John V. "Conservation Reconsidered," first published in *American Economic Review* Vol. 57 (1967).

EXAMPLE 4.2

Historical Example: Valuing the Northern Spotted Owl

The Northern Spotted Owl lives in an area of the Pacific Northwest where its habitat is threatened by logging. Its significance derives not only from its designation under the Endangered Species Act as a threatened species, but also from its role as an indicator of the overall health of the Pacific Northwest's old-growth forest.

In 1990 an interagency scientific committee presented a plan to withdraw certain forested areas from harvesting and preserve them as "habitat conservation areas." Would preserving these areas represent an efficient choice?

To answer this question, a national contingent valuation survey (this technique is outlined in the next section) was conducted to estimate the nonuse value of preservation in this case. Conducted by mail, the survey went to 1,000 households.

The results suggested that the benefits of preservation outweighed the costs by at least 3 to 1, regardless of the assumptions necessitated by the need to resolve such issues as how to treat the nonresponding households. (One calculation, for example, included them all as a zero nonuse value.) Under the assumptions most favorable to preservation, the ratio of benefits to costs was 43 to 1. In this case the nonuse values were large enough to indicate that preservation was the preferred choice.

The authors also point out, however, that the distributional implications of this choice should not be ignored. While the benefits of preservation are distributed widely throughout the entire population, the costs are concentrated on a relatively small group of people in one geographic region. Perhaps the public should be willing to share some of the preservation costs by allocating tax dollars to this area to facilitate the transition and to reduce the hardship. Ultimately this is what happened.

Since this time, multiple studies have sought to value threatened or endangered species. In a survey of the literature, Richardson and Loomis (2009) find 31 contingent valuation studies valuing threatened and endangered species based in the United States and 12 studies outside of the United States.

Sources: Daniel A. Hagen, James W. Vincent, and Patrick G. Welle, "Benefits of Preserving Old-Growth Forests and the Spotted Owl." CONTEMPORARY POLICY ISSUES, Vol. 10, April 1992, pp. 13–26; Leslie Richardson and John Loomis, "Total Economic Valuation of Endangered Species: A Summary and Comparison of the United States and the Rest of the World Estimates." CONSERVING AND VALUING ECOSYSTEM SERVICES AND BIODIVERSITY: ECONOMIC, INSTITUTIONAL AND SOCIAL CHALLENGES, K. N. Ninan, ed., Earthscan, 2009.

the U.S. Department of Interior drew up its regulations on the appropriate procedures for performing natural resource damage assessment, it prohibited the inclusion of nonuse values unless use values for the incident under consideration were zero. A subsequent 1989 decision by the District of Columbia Court of Appeals (880 F. 2nd 432) overruled this decision and allowed nonuse values to be included as long as they could be measured reliably.

Classifying Valuation Methods

Typically, the researcher's goal is to estimate the total willingness to pay for the good or service in question. This is the area under the demand curve up to the quantity consumed (recall discussion from Chapter 2). For a market good, this calculation is relatively straightforward. However, nonmarket goods and services, the focus of this chapter, require the estimation of willingness to pay either through examining behavior, drawing inferences from the demand for related goods, or through responses to surveys. And, as highlighted above, capturing all components of value is challenging.

This section will provide a brief overview of some of the methods available to estimate these values and to convey some sense of the range of possibilities and how they are related. Subsequent sections will provide more specific information about how they are actually used.

Valuation Methods. Valuation methods can be separated into two broad categories: stated preference and revealed preference methods. Each of these broad categories of methods includes both indirect and direct techniques. The possibilities are presented in Table 4.1. Revealed preference methods are those that are based on actual observable choices that allow resource values to be directly inferred from those choices. For example, in calculating how much local fishermen lost from the oil spill, the revealed preference method might calculate how much the catch declined and the resulting diminished value of the catch. In this case, prices are directly observable, and their use allows the direct calculation of the loss in value. Or, more indirectly, in calculating the value of an occupational environmental risk (such as some exposure to a substance that could pose some health risk), we might examine the differences in wages across industries in which workers take on different levels of risk.

Compare this with the direct stated preference case that might be used when the value is not directly observable. In Example 4.1, for example, the nonuse value of the Northern Spotted Owl was not directly observable. Hence, the authors attempted to derive this value by using a survey that attempted to elicit the respondents' willingness to pay (their "stated preference") for the preservation of the species.

Table 4.1 Economic Methods for Measuring Environmental and Resource Values

Methods	Revealed Preference	Stated Preference
Direct	Market Price	Contingent Valuation
	Simulated Markets	
Indirect	Travel Cost	Attribute-Based Models
	Hedonic Property Values	Conjoint Analysis
	Hedonic Wage Values	Choice Experiments
	Avoidance Expenditures	Contingent Ranking

Source: Modified by the author from Mitchell and Carson, 1989.

Stated Preference Methods

Stated preference methods use survey techniques to elicit willingness to pay for a marginal improvement or for avoiding a marginal loss. The most direct approach, called *contingent valuation*, provides a means of deriving values that cannot be obtained in more traditional ways. The simplest version of this approach merely asks respondents what value they would place on an environmental change (such as the loss of a wetlands or increased exposure to pollution) or on preserving the resource in its current state. Alternative versions ask a “yes” or “no” question such as whether or not the respondent would pay $*X* to prevent the change or preserve the species. The answers reveal either an upper bound (in the case of a “no” answer) or a lower bound (in the case of a “yes” answer). This survey approach creates a hypothetical market and asks respondents to consider a willingness-to-pay question *contingent* on the existence of this market. The major concern with the use of the contingent valuation method has been the potential for survey respondents to give biased answers. Five types of potential bias have been the focus of a large amount of research: (1) strategic bias, (2) information bias, (3) starting-point bias, (4) hypothetical bias, and (5) the observed discrepancy between willingness to pay (WTP) and willingness to accept (WTA).

Strategic bias arises when the respondent provides a biased answer in order to influence a particular outcome. If a decision to preserve a stretch of river for fishing, for example, depends on whether or not the survey produces a sufficiently large value for fishing, the respondents who enjoy fishing may be tempted to provide an answer that ensures a high value, rather than the lower value that reflects their true valuation.

Information bias may arise whenever respondents are forced to value attributes with which they have little or no experience. For example, the valuation by a recreationist of a loss in water quality in one body of water may be based on the ease of substituting recreation on another body of water. If the respondent has no experience using the second body of water, the valuation could be based on an entirely false perception. Consider another example. Visual aides have been shown to reduce uncertainty and unfamiliarity with the good or service being valued. Labao et al. (2008) found that colored photographs, as opposed to black-and-white photographs, influence respondent willingness to pay for the Philippine Eagle. The colored photographs resulted in a higher willingness to pay than black-and-white photos. Why? The authors suggest that the higher willingness to pay could be explained by photographs in color simply providing more information or by “enhancing respondents’ ability to assimilate information.” In any case, the nature of the visual aide seems important for revealing preferences.

Starting-point bias may arise in those survey instruments in which a respondent is asked to check off his or her answers from a predefined range of possibilities. How that range is defined by the designer of the survey may affect the resulting answers. A range of $0–$100 may produce a valuation by respondents different from, for example, a range of $10–$100, even if no bids are in the $0–$10 range. Ladenburg and Olsen (2008), in a study of willingness to pay to protect nature areas in Denmark from new highway development, found that the starting-point bias in their choice experiment was gender specific, with female respondents exhibiting the greatest sensitivity to the starting point.

Hypothetical bias can enter the picture because the respondent is being confronted by a contrived, rather than an actual, set of choices. Since he or she will not actually have to pay the estimated value, the respondent may treat the survey casually, providing ill-considered answers. One early survey (Hanemann, 1994) found ten studies that directly compared willingness-to-pay estimates derived from surveys with actual expenditures. Although some of the studies found that the willingness-to-pay estimates derived from surveys exceeded actual expenditures, the majority of those found that the differences were not statistically significant.[7] More recently, Ehmke, Lusk, and List (2008) tested whether hypothetical bias depends on location and/or culture. In a study based on student experiments in China, France, Indiana, Kansas, and Niger, they found significant differences in bias across locations. Given that policy-makers frequently rely on existing benefits estimates when making decisions on other locations, this finding should not be taken lightly. The strengths and weaknesses of using estimates derived in one setting to infer benefits in another, a technique known as *benefit transfer*, are discussed below.

Increasingly, environmental economists are using these types of experiments to try to determine the severity of some of these biases as well as to learn how to reduce bias. Some of these experiments are conducted in a laboratory setting, such as a computer lab or a classroom designed for this purpose. In one such experiment on voluntary provision of public goods (donations), Landry et al. (2006) found that for door-to-door interviews, an increase in physical attractiveness of the interviewer led to sizable increases in giving. Interestingly, physical attractiveness also led to increases in response rates, particularly by male households.

The final source of bias addresses observed gaps between two supposedly closely related concepts—willingness-to-pay and willingness-to-accept compensation. Respondents to contingent valuation surveys tend to report much higher values when asked for their willingness to accept compensation for a specified loss of some good or service than if asked for their willingness to pay for a specified increase of that same good or service. Economic theory suggests the two should be equal. Debate 4.1 explores some of the reasons offered for the difference.

Much experimental work has been done on contingent valuation to determine how serious a problem these biases may present. One early survey (Carson et al., 1994) had already uncovered 1,672 contingent valuation studies and of course many more have been completed since then. Are the results from these surveys reliable enough for the policy process?

Faced with the need to compute damages from oil spills, the National Oceanic and Atmospheric Administration (NOAA) convened a panel of independent economic experts (including two Nobel Prize laureates) to evaluate the use of contingent valuation methods for determining lost passive-use or nonuse values. Their report, issued on January 15, 1993 (58 FR 4602), was cautiously supportive.

[7]For a much more skeptical view of this evidence, see (Diamond and Hausman, 1994).

The committee made clear that it had several concerns with the technique. Among those concerns, the panel listed: (1) the tendency for contingent valuation willingness-to-pay estimates to seem unreasonably large; (2) the difficulty in assuring the respondents have understood and absorbed the issues in the survey; and (3) the difficulty in assuring that respondents are responding to the specific issues in the survey rather than reflecting general warm feelings about public-spiritedness, known as the "warm glow" effect.[8]

But the panel also made clear its conclusion that suitably designed surveys could eliminate or reduce these biases to acceptable levels and it provided, in an appendix, specific guidelines for determining whether a particular study was suitably designed. The panel suggested that when practitioners follow these guidelines, they

> *can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values. . . . [A well-constructed contingent valuation study] contains information that judges and juries will wish to use, in combination with other estimates, including the testimony of expert witnesses.*

Specifically, they suggested the use of referendum-type (yes/no) willingness-to-pay questions, personal interviews when possible, clear scenario descriptions, and follow-up questions.

These guidelines have been influential in shaping more recent studies. For example, Example 4.3 shares the results of a large contingent valuation survey, designed to estimate the value of preventing future spills.

The *NOAA panel* report has created an interesting dilemma. Although it has legitimized the use of contingent valuation for estimating passive-use (nonconsumptive use) and nonuse values, the panel has also set some rather rigid guidelines that reliable studies should follow. The cost of completing an "acceptable" contingent valuation study could well be so high that they will only be useful for large incidents, those for which the damages are high enough to justify their use. Yet, due to the paucity of other techniques, the failure to use contingent valuation may, by default, result in passive-use values of zero. That is not a very appealing alternative.[9]

One key to resolving this dilemma may be provided by a technique called benefit transfer. Since original studies are time consuming and expensive, benefit transfer allows the estimates for one site to be based upon estimates from other sites or benefits from an earlier time period to provide the foundation for a current estimate.

Benefit transfer methods can take one of three forms: value transfers, benefit function transfers, or meta-analysis. Sometimes the actual benefit values derived from point estimates can simply be directly transferred from one context to another; usually adjusted for differences between the study site and the policy site. Function transfer involves using a previously estimated benefit function that relates

---

[8]A more detailed description of the methodological issues and concerns with contingent valuation with respect to the actual Exxon Valdez contingent valuation survey can be found in Mitchell (2002).

[9]Whittington (2002) examines the reasons why so many contingent valuation studies in developing countries are unhelpful. Poorly designed or rapidly implemented surveys could result in costly policy mistakes on topics that are very important in the developing world. The current push for cheaper, quicker studies is risky and researchers need to be very cautious.

DEBATE 4.1

Willingness to Pay versus Willingness to Accept: Why So Different?

Many contingent valuation studies have found that respondents tend to report much higher values for questions that ask what compensation the respondent would be willing to accept (WTA) to give something up than for questions that ask for the willingness to pay (WTP) for an incremental improvement in the same good or service. Economic theory suggests that differences between WTP and WTA should be small, but experimental findings both in environmental economics and in other microeconomic studies have found large differences. Why?

Some economists have attributed the discrepancy to a psychological endowment effect; the psychological value of something you own is greater than something you do not. In other words, you would require more compensation to be as well off without it than you would be willing to pay to get that same good and as such you would be less willing to give it up (WTA > WTP) (Kahneman, Knetsch, and Thaler, 1990). This is a form of what behavioral economists call loss aversion—the psychological premise that losses are more highly valued than gains.

Others have suggested that the difference is explainable in terms of the market context. In the absence of good substitutes, large differences between WTA and WTP would be the expected outcome. In the presence of close substitutes, WTP and WTA should not be that different, but the divergence between the two measures should increase as the degree of substitution decreases (Hanemann, 1991; Shogren et al., 1994).

The characteristics of the good may matter as well. In their review of the evidence provided by experimental studies, Horowitz and McConnell (2002) find that for "ordinary goods" the difference between WTA and WTP is smaller than the ratio of WTA/WTP for public and nonmarket goods. Their results support the notion that the nature of the property rights involved are not neutral.

The moral context of the valuation may matter as well. Croson et al. (Draft 2005) show that the amount of WTA compensation estimated in a damage case increases with the culpability of the party causing the damage as long as that party is also paying for the repairs. If, however, a third party is paying, WTA is insensitive to culpability. This difference suggests that the valuation implicitly includes an amount levied in punishment for the party who caused the damage (the valuation becomes the lost value plus a sanction).

Ultimately, the choice of which concept to use in environmental valuation comes down to how the associated property right is allocated. If someone owns the right to the resource, asking how much compensation they would take to give it up is the appropriate question. If the respondent does not have the right, using WTP to estimate the value of acquiring it is the right approach. However, as Horowitz and McConnell point out, since the holders and nonholders of "rights" value them differently, the initial allocation of property rights can have strong influence on valuation decisions for environmental amenities.

Sources: R. Croson, J. J. Rachlinski, and J. Johnston, "Culpability as an Explanation of the WTA-WTP Discrepancy in Contingent Valuation" (Draft 2005); W. M. Hanemann, "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" AMERICAN ECONOMIC REVIEW, 81, 1991, pp. 635–647; J. K. Horowitz and K. E. McConnell, "A Review of WTA/WTP Studies." JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT, 44, 2002, pp. 426–447; D. Kahneman, J. Knetsch, and R. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem." JOURNAL OF POLITICAL ECONOMY, 98, 1990, pp. 1325–1348; and J. F. Shogren, Senung Y. Shin, D. J. Hayes, and J. B. Kliebenstein, "Resolving Differences in Willingness to Pay and Willingness to Accept." AMERICAN ECONOMIC REVIEW, Vol. 84 (1), 1994, pp. 255–270.

site characteristics to site values to adjust the estimates from the original study site by entering the differentiating characteristics of the policy site in order to derive newer, more site-specific values (Johnston et al., 2006). Most recently, meta-analysis has been utilized.

Meta-analysis, sometimes called the "analysis of analyses," takes empirical estimates from a sample of studies, statistically relates them to the characteristics of the studies, and asks whether the reported differences can be attributed to differences in location, subject matter, or methodology. Meta-analysis would use this cross section of contingent valuation studies as a basis for isolating and quantifying the determinants of nonuse value. Once these determinants have been isolated and related to specific policy contexts, it may be possible to transfer estimates from one context to another by finding the value consistent with the new context without incurring the time and expense of conducting new surveys each time.

Benefit transfer methods have been widely used in situations for which financial, time, or data constraints preclude original analysis. Policy-makers frequently look to previously published studies for data that could inform a prospective decision. It has the advantage of being quick and inexpensive, but the accuracy of the estimates deteriorates the further the new context deviates temporally or spatially from the context used to derive the estimates.[10] Additionally, as we noted above, for contingent valuation estimates, Ehmke, Lusk, and List (2008) find that hypothetical bias varies considerably across countries.

Benefit transfer has not escaped controversy. Johnston and Rosenberger (2010) outline some of the potential problems with the use of benefit transfer, including a lack of studies that are both of sufficiently high quality and policy relevant. Additionally, many of the published studies do not provide enough information on the attributes to allow an assessment of how they might have affected the derived value.

In response to some of these concerns, a valuation inventory database has emerged. The Environmental Valuation Reference Inventory (EVRI) is an online searchable database of empirical studies on the economic value of environmental benefits and human health effects. It was specifically developed as a tool for use in benefit transfer. The database can be accessed at http://www.evri.ca/Global/HomeAnonymous.aspx.

A final category, indirect hypothetical methods, includes several attribute-based methods. Attribute-based methods, such as choice-based, conjoint models (or, equivalently, choice experiments), are useful when project options have multiple levels of different attributes. Like contingent valuation, choice experiments are also survey based, but instead of asking respondents to state a willingness to pay, respondents are asked to choose among alternate bundles of goods. Each bundle has a set of attributes and the levels of each attribute vary across bundles. Since one of the attributes in each bundle is a price measure, willingness to pay can be identified.

Consider an example (Boyle et al., 2001) that surveyed Maine residents on their preferences for alternative forest-harvesting practices. The State of Maine was

[10]Several examples of the use of meta-analysis and benefit transfer are given in Florax et al. (2002). A critique and alternative to benefits transfer is offered in Smith et al. (2002).

considering purchasing a 23,000 acre tract of forest land to manage. Attributes used in the survey included the number of live trees, management practice for dead trees, percent of land set aside, and a tax payment. Three levels of each management attribute and 13 different tax prices were considered. Table 4.2 reproduces the attributes and levels.

Respondents were given a choice set of four different alternative management plans and the status quo (no purchase). Table 4.3 demonstrates a sample survey question. This type of survey has evolved from both contingent valuation and marketing studies. This approach allows the respondent to make a familiar choice (choose a bundle) and allows the researcher to derive marginal willingness to pay for an attribute from that choice.

TABLE 4.2 Attributes in the Maine Forest Harvesting Conjoint Analysis

Attribute	Level
Live Trees After Harvesting	No trees (clear-cut)
	153 trees/acre
	459 trees/acre
Dead Trees After Harvesting	Remove all
	5 trees/acre
	10 trees/acre
Percent of Forest Set Aside from Harvest	20%
	50%
	80%

Source: Boyle et al., 2001 and Holmes and Adamovicz, 2003.

TABLE 4.3 A Sample Conjoint Analysis Survey Questionnaire

Attribute	Alternatives				
	A	B	C	D	No change
Live Trees Remaining	No trees	459/acre	No trees	153/acre	
Dead Trees Remaining	Remove all	Remove all	5/acre	10/acre	
Percent Set Aside	80%	20%	50%	20%	
Tax	$40	$200	$10	$80	
I would vote for (please check off)	—	—	—	—	—

Source: Thomas P. Holmes and Wiktor L. Adamowicz, "Attribute-Based Methods." A PRIMER ON NONMARKET VALUATION, Ian Bateman, ed. (New York: Kluwer Academic Publishers, 2003).

EXAMPLE 4.3

Leave No Behavioral Trace: Using the Contingent Valuation Method to Measure Passive-Use Values

Until the *Exxon Valdez* tanker spilled 11 million gallons of crude oil into Prince William Sound in Alaska, the calculation of nonuse (or passive-use) values was not a widely researched topic. However, following the 1989 court ruling in *Ohio v. U.S. Department of the Interior* that said lost passive-use values could now be compensated within natural resources damages assessments and the passage of the Oil Pollution Act of 1990, the estimation of nonuse and passive-use values became not only a topic of great debate, but also a rapidly growing research area within the economics community.

One study (Carson et al., 2003) discusses the design, implementation, and results of a large survey designed to estimate the passive-use values related to large oil spills. In particular, the survey asked respondents their willingness to pay to prevent a similar disaster in the future by funding an escort ship program that would help prevent and/or contain a future spill. The survey was conducted for the State of Alaska in preparation for litigation in the case against *Exxon Valdez.*

The survey followed the recommendations made by the NOAA panel for conducting contingent valuation surveys and for ensuring reliable estimates. It relied upon face-to-face interviews and the sample was drawn from the national population. The study used a binary discrete-choice (yes or no) question where the respondent was asked whether he or she would be willing to pay a specific amount, with the amount varying across four versions of the survey. A one-time increase in taxes was the chosen method of payment. They also avoided potential embedding bias (where respondents may have difficulty valuing multiple goods) by using a survey that valued a single good. The survey also contained pictures, maps, and background information to make sure the respondent was familiar with the good he/she was being asked to value.

Using the survey data, the researchers were able, statistically, to estimate a valuation function by relating the respondent's willingness to pay to respondent characteristics. After multiplying the estimate of the median willingness to pay by the population sampled, they reported aggregate lost passive-use values at $2.8 billion (in 1990 dollars). They point out that this number is a lower bound, not only because willingness-to-accept compensation would be a more appropriate measure of actual lost passive use from the spill (see Debate 4.1), but also because their median willingness to pay was less than the mean.

The *Exxon Valdez* spill sparked a debate about the measurement of nonuse and passive-use values. Laws put into place after the spill have ensured that passive-use values will be included in natural resource damage assessments. Should other parts of the world follow suit?

Source: Richard T. Carson, Robert C. Mitchell, Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. "Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill." ENVIRONMENTAL AND RESOURCE ECONOMICS, Vol. 25, 2003, pp. 257–286.

Contingent ranking, another survey method, also falls within this final category. Respondents are given a set of hypothetical situations that differ in terms of the environmental amenity available (instead of a bundle of attributes) and are asked to rank-order them. These rankings can then be compared to see the implicit trade-offs between more of the environmental amenity and less of the other characteristics. When one or more of these characteristics is expressed in terms of a monetary value, it is possible to use this information and the rankings to impute a value to the environmental amenity.

Sometimes more than one of these techniques may be used simultaneously. In some cases using multiple techniques is necessary to capture the total economic value; in other cases it may be used to provide independent estimates of the value being sought as a check on the reliability of the estimate.

Revealed Preference Methods

Revealed preference methods are "observable" because they involve actual behavior and "indirect" because they infer a value rather than estimate it directly. Suppose, for example, a particular sport fishery is being threatened by pollution, and one of the damages caused by that pollution is a reduction in sportfishing. How is this loss to be valued when access to the fishery is free?

Travel Cost Method

One way to derive this loss is through *travel cost* methods. Travel cost methods may infer the value of a recreational resource (such as a sport fishery, a park, or a wildlife preserve where visitors hunt with a camera) by using information on how much the visitors spent in getting to the site to construct a demand curve for willingness to pay for a "visitor day."

Freeman (2003) identifies two variants of this approach. In the first, analysts examine the number of trips visitors make to a site. In the second, the analysts examine whether people decide to visit a site and, if so, which site. This second variant includes using a special class of models, known as random utility models, to value quality changes.

The first variant allows the construction of a travel cost demand function. The value of the flow of services from that site is the area under the estimated demand curve for those services or for access to the site, aggregated over all who visit the site.

The second variant allows the analysis of how specific site characteristics influence choice and, therefore, indirectly how valuable those characteristics are. Knowledge of how the value of each site varies with respect to its characteristics allows the analyst to value how degradation of those characteristics (e.g., from pollution) would lower the value of the site.

Travel cost models have been used to value beach closures during oil spills, fish consumption advisories, and the cost of development that has eliminated a recreation area. The methodology for both variants is detailed in (Parsons, 2003).

In the random utility model, a person choosing a particular site takes into consideration site characteristics and its price (trip cost). Characteristics affecting the site choice include ease of access and environmental quality. Each site results in a unique level of utility and a person is assumed to choose the site giving the highest level of utility to that person. Welfare losses from an event such as an oil spill can then be measured by the resulting change in utility should the person have to choose an alternate, less desirable site.

One interesting paradox that arises with the travel-cost model is that those who live closest to the site and may actually visit frequently, will have low travel costs. These users will appear to have a lower value for that site even if their (unmeasured) willingness to pay for the experience is very high. Another challenge in this model is how to incorporate the opportunity cost of time. Usually, this is represented by wages, but that approach is not universally accepted.

Hedonic Property Value and Hedonic Wage Methods

Two other indirect observable methods are known as *the hedonic property value* and *hedonic wage* approaches. They share the characteristic that they use a statistical technique, known as multiple regression analysis, to "tease out" the environmental component of value in a related market. For example, it is possible to discover that, all other things being equal, property values are lower in polluted neighborhoods than in clean neighborhoods. (Property values fall in polluted neighborhoods because they are less desirable places to live.)

Hedonic property value models use market data (house prices) and then break down the house sales price into its components, including the house characteristics (e.g., number of bedrooms, lot size, and features); the neighborhood characteristics (e.g., crime rates, school quality, and so on); and environmental characteristics (e.g., air quality, percentage of open space nearby, and distance to a local landfill). Hedonic models allow for the measurement of the marginal willingness to pay for discrete changes in an attribute. Numerous studies have utilized this approach to examine the effect on property value of things such as distance to a hazardous waste site (Michaels and Smith, 1990); large farm operations (Palmquist et al., 1997); and open space and land use patterns (Bockstael, 1996; Geoghegan et al., 1997; Acharya and Bennett, 2001).

Hedonic wage approaches are similar except that they attempt to isolate the environmental risk component of wages, which serves to isolate the amount of compensation workers require in order to work in risky occupations. It is well known that workers in high-risk occupations demand higher wages in order to be induced to undertake the risks. When the risk is environmental (such as exposure to a toxic substance), the results of the multiple regression analysis can be used to construct a willingness to pay to avoid this kind of environmental risk. Additionally, the compensating wage differential can be used to calculate the value of a statistical life (Taylor, 2003). Techniques for valuing reductions in life-threatening risks will be discussed later in this chapter.

Averting Expenditures

A final example of an indirect observable method involves examining "averting or defensive expenditures." Averting expenditures are those designed to reduce the damage caused by pollution by taking some kind of averting or defensive action. An example would be to install indoor air purifiers in response to an influx of polluted air or to rely on bottled water as a response to the pollution of local drinking water supplies (Example 4.4). Since people would not normally spend more to prevent a problem than would be caused by the problem itself, averting expenditures can provide a lower-bound estimate of the damage caused by pollution.

Using Geographic Information Systems for Economic Valuation

Geographic Information Systems (GIS) are computerized mapping models and analysis tools. A GIS map is made up of layers such that many variables can be visualized simultaneously using overlays. Use of GIS to inform economic analysis is

EXAMPLE 4.4

Valuing Damage from Groundwater Contamination Using Averting Expenditures

How many resources should be allocated to the prevention of groundwater contamination? In part, that depends on how serious a risk is posed by the contamination. How much damage would be caused? One way to obtain a lower-bound estimate on the damage caused is to discover how much people are willing to spend to defend themselves against the threat.

In late 1987, trichloroethylene (TCE) was detected in one of the town wells in Perkasie, a town in southeastern Pennsylvania. Concentrations of the chemical were seven times the EPA's safety standard. Since no temporary solution was available to reduce concentrations to safe levels, the county required the town to notify customers of the contamination.

Once notified, consumers took one or more of the following actions: (1) they purchased more bottled water; (2) they started using bottled water; (3) they installed home water-treatment systems; (4) they hauled water from alternative sources; and (5) they boiled water. Through a survey, analysts were able to discover the extent of each of these actions and combine that information with their associated costs.

The results indicated that residents spent between \$61,313.29 and \$131,334.06 over the 88-week period of the contamination to protect themselves from the effects. They further indicated that families with young children were more likely to take averting actions and, among those families who took averting actions, to spend more on those actions than childless families.

Source: Charles W. Abdalla et al., "Valuing Environmental Quality Changes Using Averting Expenditures: An Application to Groundwater Contamination." LAND ECONOMICS, Vol. 68, No. 2 (1992), pp. 163–169.

a relatively recent addition to the economist's tool kit. GIS offers a powerful collection of tools for depicting and examining spatial relationships. Most simply, GIS can be used to produce compelling graphics that communicate the spatial structure of data and analytic results with a force and clarity otherwise impossible. But the technology's real value lies in the potential it brings to ask novel questions and enrich our understanding of social and economic processes by explicitly considering their spatial structure. Models that address environmental externalities have, almost by definition, a strong spatial component. One study (Bateman et al., 2002) examines the contributions of GIS in incorporating spatial dimensions into economic analysis, including benefit–cost analysis. Another study (Clapp et al., 1997) discusses the potential contributions GIS can make for urban and real estate economics.

Hedonic property valuation models have recently incorporated GIS technology. Fundamentally spatial in nature, use of GIS by hedonic property models is a natural fit. Housing prices vary systematically and predictably from neighborhood to neighborhood. Spatial characteristics, from air quality to the availability of open space, can influence property values of entire neighborhoods; if one house enjoys abundant open space or especially good air quality, it is highly likely that its neighbors do as well.

In a 2008 paper, Lewis, Bohlen, and Wilson use GIS and statistical analysis to evaluate the impacts of dams and dam removal on local property values. In a unique "experiment" they collected data on property sales for ten years before and after the Edwards Dam on the Kennebec River in Maine was removed. The Edwards Dam was the first federally licensed hydropower dam in the United States to be removed primarily for the purpose of river restoration. They also collected data on property sales approximately 20 miles upstream where two dams were still in place. GIS technology enhanced this study by facilitating the calculation of the distance from each home to both the river and the nearby dams. Lewis et al. found that homeowners pay a price penalty for living close to a dam. In other words, willingness to pay for identical housing is higher, the further away from the dam the house is located. They also found that the penalty near the Edwards Dam site dropped to nearly zero after the dam was removed. Interestingly, the penalty upstream also dropped significantly. While a penalty for homes close to the dams upstream remains, it falls after the downstream dam was removed. Can you think of reasons why?[11]

Example 4.5 shows how the use of GIS can enable hedonic property value models to investigate how the view from a particular piece of property might affect its value.

Valuing Human Life. One fascinating public policy area where these various approaches have been applied is in the valuation of human life. Many government programs, from those controlling hazardous pollutants in the workplace or in

[11] Interestingly, after this study was complete, one of the two upstream dams, the Fort Halifax Dam, was removed in July 2008 after years of litigation about its removal.

EXAMPLE 4.5

Using GIS to Inform Hedonic Property Values: Visualizing the Data

Geographic information systems (GIS) offer economists and others powerful tools for analyzing spatial data and spatial relationships. For nonmarket valuation, GIS has proven to be especially helpful in enhancing hedonic property value models by incorporating both the proximity of environmental characteristics and their size or amount. GIS studies have also allowed for the incorporation of variables that reflect nearby types and diversity of land use.

Geo-coding housing transactions assigns a latitude and longitude coordinate to each sale. GIS allows other spatial data, such as land use, watercourses, and census data, to be "layered" on top of the map. By drawing a circle around each house of the desired circumference, GIS can help us to calculate the amount of each amenity that is in that circle as well as the density and types of people who live there. Numerous census data are available on variables such as income, age, education, crime rates, and commuting time. GIS also makes it relatively easy to calculate straight-line distances to desired (or undesired) locations, such as parks, lakes, schools, or landfills.

In a 2002 paper entitled "Out of Sight, Out of Mind? Using GIS to Incorporate Visibility in Hedonic Property Value Models," Paterson and Boyle use GIS to measure the extent to which visibility measures affect house prices in Connecticut. In their study visibility is measured as the percentage of land visible within one kilometer of the property, both in total and broken out for various land use categories. Finally, they added variables that measured the percentage of area in agriculture or in forest, or covered by water within one kilometer of each house.

They find that visibility is indeed an important environmental variable in explaining property values, but the nature of the viewshed matters. While simply having a view is not a significant determinant of property values, viewing certain types of land uses is. Proximity to development reduces property values only if the development is visible, for example, suggesting that out of sight, really does mean out of mind! They conclude that any analysis that omits variables that reflect nearby environmental conditions can lead to misleading or incorrect conclusions about the impacts of land use on property values. GIS is a powerful tool for helping a researcher include these important variables.

Source: Robert Paterson and Kevin Boyle, "Out of Sight, Out of Mind? Using GIS to Incorporate Visibility in Hedonic Property Value Models." LAND AND ECONOMICS, 78(3), 2002, pp.417–425.

drinking water, to those improving nuclear power plant safety, are designed to save human life as well as to reduce illness. How resources should be allocated among these programs depends crucially on the value of human life. In order to answer this question, an estimate of the value of that life to society is necessary and federal regulations require such estimates for benefit–cost analysis. How is life to be valued?

The simple answer, of course, is that life is priceless, but that turns out to be not very helpful. Because the resources used to prevent loss of life are scarce, choices must be made. The economic approach to valuing lifesaving reductions in environmental risk is to calculate the change in the probability of death resulting from the reduction in environmental risk and to place a value on the change. Thus, it is not life itself that is being valued, but rather a reduction in the probability that some segment of the population could be expected to die earlier than otherwise. This *value of statistical life* (VSL) represents an individual's willingness to pay for small changes in mortality risks. It does not represent a willingness to pay to prevent certain death. It is measured as the "marginal rate of substitution between mortality risk and money (i.e., other goods and services)" (Cameron 2011). Debate 4.2 examines the controversy associated with valuing changes in these mortality risks.

DEBATE 4.2

Is Valuing Human Life Immoral?

In 2004 economist Frank Ackerman and lawyer Lisa Heinzerling teamed up to write a book that questions the morality of using benefit–cost analysis to evaluate regulations designed to protect human life. In *Priceless: On Knowing the Price of Everything and the Value of Nothing* (2004), they argue that benefit–cost analysis is immoral because it represents a retreat from the traditional standard that all citizens have an absolute right to be free from harm caused by pollution. When it justifies a regulation that will allow some pollution-induced deaths, benefit–cost analysis violates this absolute right.

Economist Maureen Cropper responds that it would be immoral not to consider the benefits of lifesaving measures. Resources are scarce and they must be allocated so as to produce the greatest good. If all pollution were reduced to zero, even if that were possible, the cost would be extremely high and the resources to cover that cost would have to be diverted from other beneficial uses. Professor Cropper also suggests that it would be immoral to impose costs on people about which they have no say—for example, the costs of additional pollution controls—without at least trying to consider what choices people would make themselves. Like it or not, hard choices must be made.

Cropper also points out that people are always making decisions that recognize a trade-off between the cost of more protection and the health consequences of not taking the protection. Thinking in terms of trade-offs should be a familiar concept. She points out that people drive faster to save time, thereby increasing their risk of dying. They also decide how much money to spend on medicines to lower their risk of disease or they may take jobs that pose morbidity or even mortality risks.

In her response to Ackerman and Heinzerling, Cropper acknowledges that benefit–cost analysis has its flaws and that it should never be the only decision-making guide. Nonetheless, she argues that it does add useful information to the process and throwing that information away could prove to be detrimental to the very people that Ackerman and Heinzerling seek to protect.

Sources: Frank Ackerman and Lisa Heinzerling, PRICELESS: ON KNOWING THE PRICE OF EVERYTHING AND THE VALUE OF NOTHING (New York: The New Press, 2004); Frank Ackerman, "Morality, Cost-Benefit and the Price of Life." ENVIRONMENTAL FORUM, Vol. 21, No. 5 (2004), pp. 46–47; and Maureen Cropper, "Immoral Not to Weigh Benefits Against Costs." ENVIRONMENTAL FORUM, 21, No. 5 (2004): 47–48.

It is possible to translate the value derived from this procedure into an "implied value of human life." This is accomplished by dividing the amount each individual is willing to pay for a specific reduction in the probability of death by the probability reduction. Suppose, for example, that a particular environmental policy could be expected to reduce the average concentration of a toxic substance to which one million people are exposed. Suppose further that this reduction in exposure could be expected to reduce the risk of death from 1 out of 100,000 to 1 out of 150,000. This implies that the number of expected deaths would fall from 10 to 6.67 in the exposed population as a result of this policy. If each of the one million persons exposed is willing to pay $5 for this risk reduction (for a total of $5 million), then the implied value of a statistical life is approximately $1.5 million ($5 million divided by 3.33). Or alternatively, the VSL can be calculated using the change in WTP divided by the change in risk. For this example, that would be $5 divided by the change in risk of death (1/100,000–1/150,000), or $1.5 million. Thus, the VSL is capturing the rate of trade-off between money and a very small risk of death.

What actual values have been derived from these methods? One early survey (Viscusi, 1996) of a large number of studies examining reductions in a number of life-threatening risks found that most implied values for human life (in 1986 dollars) were between $3 million and $7 million. This same survey went on to suggest that the most appropriate estimates were probably closer to the $5 million estimate. In other words, all government programs resulting in risk reductions costing less than $5 million per life saved would be justified in benefit–cost terms. Those costing more might or might not be justified, depending on the appropriate value of a life saved in the particular risk context being examined.

In a recent meta-analysis, Mrozek and Taylor (2002) found much lower values for VSL. Using over 40 labor market studies, their research suggest that a range of $1.5 million to $2.5 million for VSL is more appropriate. What about age? Does the VSL change with age? Apparently so. Aldy and Viscusi (2008) find an inverted U-shape relationship between VSL and age. Specifically, using hedonic wage model, they estimate a VSL of $3.7 million for persons ages 18–24, $9.7 million for persons ages 35–44, and $3.4 million for persons ages 55–62. VSL rises with age, peaks, and then declines.

What about the value of statistical life across populations or countries with different incomes? How does VSL vary with income? Most agencies in the United States use VSLs between $5 million and $8 million. These estimates are based largely on hedonic wage studies that have been conducted in the United States or in other high-income countries.[12] How might those results be translated into settings featuring populations with lower incomes?

Adjustments for income are typically derived using an estimate of the income elasticity of demand. Recall that income elasticity is the percent change in consumption given a 1 percent change in income. Hammitt and Robinson (2011) note that applying income elasticities, derived for countries like the United States, might result in nonsensical VSL estimates if blindly applied to lower-income countries. While U.S. agencies typically assume a 0.4 to 0.6 percent change in VSL for a 1 percent change in real income over time, elasticities closer to 1.0 or higher are more realistic for

[12]Many labor market estimates of VSL average near $7 million (Viscusi 2008).

transfers of these values between high- and low-income countries. Using the higher income elasticity number is merited since willingness to pay for mortality risk reduction as a percentage of income drops at very low incomes; what limited income is available in poorer households is reserved for basic needs.

How have health, safety, and environmental regulations lived up to this recommendation? As Table 4.4 suggests, not very well. A very large number of regulations listed in that table could be justified only if the value of a life saved were much higher than the upper value of $7 million.

TABLE 4.4 The Cost of Risk-Reducing Regulations

	Agency Year and Status	Initial Annual Risk	Annual Lives Saved	Cost Per Life Saved (Millions of 1984 $)
Unvented Space Heaters	CPSC 1980 F	2.7 in 10^5	63.000	$.10
Cabin Fire Protection	FAA 1985 F	6.5 in 10^8	15.000	.20
Passive Restraints/Belts	NHTSA 1984 F	9.1 in 10^5	1,850.000	.30
Seat Cushion Flammability	FAA 1984 F	1.6 in 10^7	37.000	.60
Floor Emergency Lighting	FAA 1984 F	2.2 in 10^8	5.000	.70
Concrete and Masonry Construction	OSHA 1988 F	1.4 in 10^5	6.500	1.40
Hazard Communication	OSHA 1983 F	4.0 in 10^5	200.000	1.80
Benzene/Fugitive Emissions	EPA 1984 F	2.1 in 10^4	0.310	2.80
Radionuclides/Uranium Mines	EPA 1984 F	1.4 in 10^4	1.100	6.90
Benzene	OSHA 1987 F	8.8 in 10^4	3.800	17.10

(continued)

(Continued)

	Agency Year and Status	Initial Annual Risk	Annual Lives Saved	Cost Per Life Saved (Millions of 1984 $)
Asbestos	EPA 1989 F	2.9 in 10^5	10.000	104.20
Benzene/Storage	EPA 1984 R	6.0 in 10^7	0.043	202.00
Radionuclides/ DOE Facilities	EPA 1984 R	4.3 in 10^6	0.001	210.00
Radionuclides/ Elemental Phosphorous	EPA 1984 R	1.4 in 10^5	0.046	270.00
Benzene/ Ethylbenzenol Styrene	EPA 1984 R	2.0 in 10^6	0.006	483.00
Arsenic/ Low-Arsenic Copper	EPA 1986 R	2.6 in 10^4	0.090	764.00
Benzene/ Maleic Anhydride	EPA 1984 R	1.1 in 10^6	0.029	820.00
Land Disposal	EPA 1988 F	2.3 in 10^8	2.520	3,500.00
Formaldehyde	OSHA 1987 F	6.8 in 10^4	0.010	72,000.00

Note: In the "Agency Year and Status" column, R and F represent Rejected and Final rule, respectively. "Initial Annual Risk" indicates annual deaths per exposed population; an exposed population of 10^3 is 1000, 10^4 is 10,000, and so on.

Source: Data from Tables 1 and 2 from "Economic Foundation of the Current Regulatory Reform Efforts" by W. Kip Viscusi, from JOURNAL OF ECONOMIC PERSPECTIVES, 10 (3) summer, 1996, pp. 119–134.

Summary: Nonmarket Valuation Today

In this chapter we have examined the most prominent, but certainly not the only techniques available to supply policy-makers with the information needed to implement efficient policy. Finding the total economic value of the service flows requires estimating three components of value: (1) use value, (2) option value, and (3) nonuse or passive-use value.

Our review of these various techniques included direct observation, contingent valuation, contingent choice experiments, travel cost, hedonic property and wage studies, and averting or defensive expenditures. When time or funding precludes original research, benefits transfer or meta-analysis provides alternate methods for estimation of values. Examples of actual studies using these techniques were presented.

In January 2011, a panel of experts gathered at the annual meeting of the American Economics Association to reflect on nonmarket valuation 20 years after the *Exxon Valdez* spill and, unknown to any of them when the panelists were asked to participate, eight months after the *Deepwater Horizon* spill. The panelists had all worked on estimation of damages from the *Exxon Valdez* spill and included Kevin Boyle, Richard Carson, Joseph Herriges, Ted McConnell, and V. Kerry Smith. The consensus among panelists was that while many of the issues with bias have been addressed in the literature, many unanswered questions remain and some areas still need work. While they all agreed that it is "hard to underestimate the powerful need for values" (i.e., some number is definitely better than no number), and we now have in place methods that can be easily utilized by all researchers, they also emphasized several problem areas. First, the value of time in travel cost models has not been resolved. What is the opportunity cost of time if you are unemployed, for example? In discussing other revealed preference methods, they asked the question, "How do the recent numerous foreclosures in the real estate market affect hedonic property value model assumptions?" Second, choice experiments do not resolve all of the potential problems with contingent valuation. While choice experiments do seem to better represent actual market choices, some of the issues that arise in contingent valuation, such as the choice of the payment vehicle, also arise with choice experiments. In addition, some new challenges, such as how the sequencing of choices in choice experiments might affect outcomes, arise. The panel highlighted how this area of research has been enhanced by the field of behavioral economics, an emerging research area that combines economics and psychology to examine human behavior. And finally, they suggested that the NOAA panel recommendations be updated to reflect the body of new research.

Discussion Questions

1. Certain environmental laws prohibit EPA from considering the costs of meeting various standards when the levels of the standards are set. Is this a good example of "putting first things first" or simply an unjustifiable waste of resources? Why?

Self-Test Exercises

1. In Mark A. Cohen, "The Costs and Benefits of Oil Spill Prevention and Enforcement," *Journal of Environmental Economics and Management* Vol. 13 (June 1986), an attempt was made to quantify the marginal benefits and marginal costs of U.S. Coast Guard enforcement activity in the area of oil

spill prevention. His analysis suggests (p. 185) that the marginal per-gallon benefit from the current level of enforcement activity is $7.50, while the marginal per-gallon cost is $5.50. Assuming these numbers are correct, would you recommend that the Coast Guard increase, decrease, or hold at the current level their enforcement activity? Why?

2. In Table 4.4, Professor Kip Viscusi estimates that the cost per life saved by current government risk-reducing programs ranges from $100,000 for unvented space heaters to $72 billion for a proposed standard to reduce occupational exposure to formaldehyde.
   a. Assuming these values to be correct, how might efficiency be enhanced in these two programs?
   b. Should the government strive to equalize the marginal costs per life saved across all lifesaving programs?
3. a. Suppose that hedonic wage studies indicate a willingness to pay $50 per person for a reduction in the risk of a premature death from an environmental hazard of 1/100,000. If the exposed population is four million people, what is the implied value of a statistical life?
   b. Suppose that an impending environmental regulation to control that hazard is expected to reduce the risk of premature death from 6/100,000 to 2/100,000 per year in that exposed population of four million people. Your boss asks you to tell her what is the maximum this regulation could cost and still have the benefits be at least as large as the costs. What is your answer?

Further Reading

Bateman, Ian J., Andrew A. Lovett, and Julii S. Brainard. *Applied Environmental Economics: A GIS Approach to Cost-Benefit Analysis* (Cambridge: Cambridge University Press, 2005). Uses GIS to examine land use change and valuation.

Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer. *Cost-Benefit Analysis: Concepts and Practice*, 3rd ed. (Upper Saddle River, NJ: Prentice-Hall, 2005). An excellent basic text on the use of cost-benefit analysis.

Champ, Patricia A., Kevin J. Boyle, and T. C. Brown. *A Primer on Nonmarket Valuation*, (Dordrecht: Kluwer Academic Publishers, 2003). A thorough overview of nonmarket valuation methods.

Costanza, R. et al. "The Value of the World's Ecosystem Services and Natural Capital" (Reprinted from *Nature* Vol. 387, p. 253, 1997), *Ecological Economics*, Vol. 25, No. 1 (1998): 3–15. An ambitious, but ultimately flawed, attempt to place an economic value on ecosystem services. This issue of *Ecological Economics* also contains a number of articles that demonstrate some of the flaws.

Hausman, Jerry A., ed. *Contingent Valuation: A Critical Assessment* (Amsterdam: North Holland, 1993). The critics of contingent valuation weigh in. Kopp, Raymond J., and V. Kerry Smith, eds. *Valuing Natural Assets: The Economics of Natural Resource Damage Assessment* (Washington, DC: Resources for the Future, Inc., 1993). A comprehensive set

of essays by some of the chief practitioners in the field evaluating both the legal framework for damage assessment and the validity and reliability of the methods currently being used. Robert Cameron, Mitchell, and Richard T. Carson. *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Washington, DC: Resources for the Future, 1989). A comprehensive examination of contingent valuation research with brief summaries of representative studies.

Lewis, Lynne Y., Curtis Bohlen, and Sarah Wilson. "Dams, Dam Removal and River Restoration: A Hedonic Analysis," *Contemporary Economic Policy* Vol. 26, No. 2 (2008): 175–186.

Additional References and Historically Significant References are available on this book's Companion Website: http://www.pearsonhighered.com/tietenberg/

5 Dynamic Efficiency and Sustainable Development

We usually see only the things we are looking for—so much so that we sometimes see them where they are not

—Eric Hoffer, *The Passionate State of Mind* (1993)

Introduction

In previous chapters, we have developed two specific means for identifying environmental problems. The first, static efficiency, allows us to evaluate those circumstances where time is not a crucial aspect of the allocation problem. Typical examples might include allocating resources such as water or solar energy where next year's flow is independent of this year's choices. The second, more complicated criterion, dynamic efficiency, is suitable for those circumstances where time is a crucial aspect. One typical example might include the combustion of depletable resources such as oil, since supplies used now are unavailable for future generations.

After defining these criteria and showing how they could be operationally invoked, we demonstrated how helpful they can be. They are useful not only in identifying environmental problems and ferreting out their behavioral sources, but also in providing a basis for identifying types of remedies. These criteria even help design the various policy instruments that can be used to restore some sense of balance.

But the fact that these are powerful and useful tools in the quest for a sense of harmony between the economy and the environment does not imply that they are the only criteria in which we should be interested. In a general sense, the efficiency criteria are designed to prevent wasteful use of environmental and natural resources. That is a desirable attribute, but it is not the only possible desirable attribute. We might care, for example, not only about the value of the environment (the size of the pie), but also how this value is shared (the size of each piece to all recipients). In other words, fairness or justice concerns should accompany efficiency considerations.

In this chapter, we investigate one particular fairness concern—the treatment of future generations. We begin by considering a specific, ethically challenging situation—the allocation of a depletable resource over time. Using a numerical example, we shall trace out the temporal allocation of a depletable resource using the dynamic efficiency criterion and show how this allocation is affected by changes in the discount rate. To lay the groundwork for our evaluation of fairness, we then turn to the task of defining what we mean by intertemporal fairness. Finally, we consider not only how this theoretical definition can be made operationally measurable, but also how it relates to dynamic efficiency. To what degree is dynamic efficiency compatible with intergenerational fairness?

A Two-Period Model

Dynamic efficiency balances present and future uses of a depletable resource by maximizing the present value of the net benefits derived from its use. This implies a particular allocation of the resource across time. We can investigate the properties of this allocation and the influence of such key parameters as the discount rate with the aid of a simple numerical example. We begin with the simplest of models—deriving the dynamic efficient allocation across two time periods. In subsequent chapters, we show how these conclusions generalize to longer time periods and to more complicated situations.

Assume that we have a fixed supply of a depletable resource to allocate between two periods. Assume further that the demand function is constant in the two periods, the marginal willingness to pay is given by the formula $P = 8 - 0.4q$, and marginal cost is constant at \$2 per unit (see Figure 5.1). Note that if the total supply was 30 or greater, and we were concerned only with these two periods, an efficient allocation would produce 15 units in each period, *regardless of the discount rate*.

FIGURE 5.1 The Allocation of an Abundant Depletable Resource: (a) Period 1 and (b) Period 2

Thirty units would be sufficient to cover the demand in both periods; the consumption in Period 1 does not reduce the consumption in Period 2. In this case the static efficiency criterion is sufficient because the allocations are not interdependent.

Examine, however, what happens when the available supply is less than 30. Suppose it equals 20. How do we determine the efficient allocation? According to the dynamic efficiency criterion, the efficient allocation is the one that maximizes the present value of the net benefit. The present value of the net benefit for both periods is simply the sum of the present values in each of the two periods. To take a concrete example, consider the present value of a particular allocation: 15 units in the first period and 5 in the second. How would we compute the present value of that allocation?

The present value in the first period would be that portion of the geometric area under the demand curve that is over the supply curve—$45.00.[1] The present value in the second period is that portion of the area under the demand curve that is over the supply curve from the origin to the five units produced multiplied by $1/(1 + r)$. If we use $r = 0.10$, then the present value of the net benefit received in the second period is $22.73,[2] and the present value of the net benefits for the two years is $67.73.

Having learned how to find the present value of net benefits for any allocation, how does one find the allocation that maximizes present value? One way, with the aid of a computer, is to try all possible combinations of q_1 and q_2 that sum to 20. The one yielding the maximum present value of net benefits can then be selected. That is tedious and, for those who have the requisite mathematics, unnecessary.

The dynamically efficient allocation of this resource has to satisfy the condition that the present value of the marginal net benefit from the last unit in Period 1 equals the present value of the marginal net benefit in Period 2 (see appendix at the end of this chapter). Even without mathematics, this principle is easy to understand, as can be demonstrated with the use of a simple graphical representation of the two-period allocation problem.

Figure 5.2 depicts the present value of the marginal net benefit for each of the two periods. The net benefit curve for Period 1 is to be read from left to right. The net benefit curve intersects the vertical axis at $6; demand would be zero at $8 and the marginal cost is $2, so the difference (marginal net benefit) is $6. The marginal net benefit for the first period goes to zero at 15 units because, at that quantity, the willingness to pay for that unit exactly equals its cost.

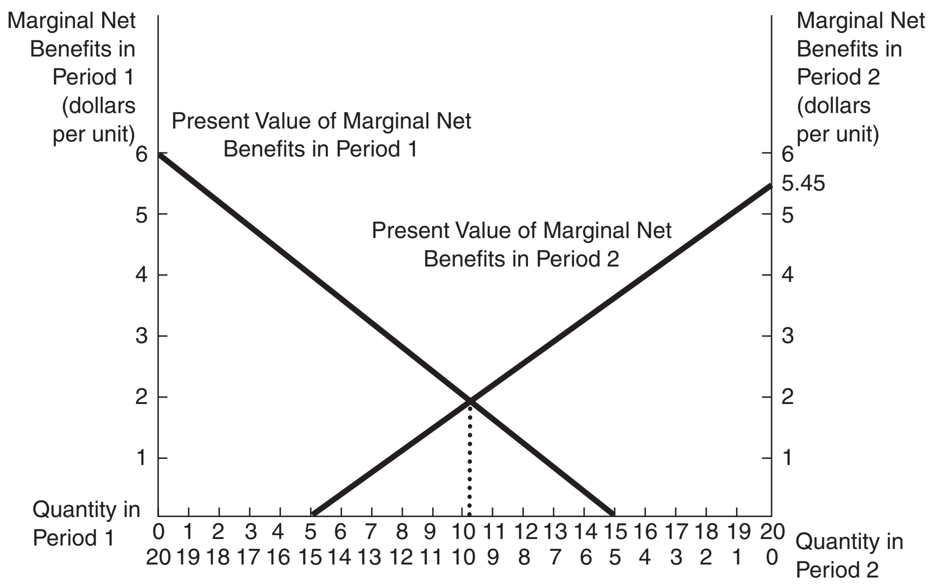
The only challenging aspect of drawing the graph involves constructing the curve for the present value of net benefits in Period 2. Two aspects are worth noting. First, the zero axis for the Period 2 net benefits is on the right, rather than the left, side. Therefore, increases in Period 2 are recorded from right to left. This way, all points along the horizontal axis yields a total of 20 units

---

[1]The height of the triangle is $6 [$8–$2] and the base is 15 units. The area is therefore (1/2)($6)(15) = $45.

[2]The undiscounted net benefit is $25. The calculation is $(6 - 2) \times 5 + 1/2(8 - 6) \times 5 = \$25$. The discounted net benefit is therefore 25/1.10 = 22.73.

FIGURE 5.2 The Dynamically Efficient Allocation



allocated between the two periods. Any point on that axis picks a unique allocation between the two periods.[3]

Second, the present value of the marginal benefit curve for Period 2 intersects the vertical axis at a different point than does the comparable curve in Period 1. (Why?) This intersection is lower because the marginal benefits in the second period need to be discounted (multiplied by $1/(1 + r)$) to convert them into present value form since they occur one year later. Thus, with the 10 percent discount rate we are using, the marginal net benefit is \$6 and the present value is \$6/1.10 = \$5.45. Note that larger discount rates would rotate the Period 2 marginal benefit curve around the point of zero net benefit ($q_1 = 5$, $q_2 = 15$) toward the right-hand axis. We shall use this fact in a moment.

The efficient allocation is now readily identifiable as the point where the two curves representing present value of marginal net benefits cross. The total present value of net benefits is then the area under the marginal net benefit curve for Period 1 up to the efficient allocation, plus the area under the present value of the marginal net benefit curve for Period 2 from the right-hand axis up to its efficient allocation. Because we have an efficient allocation, the sum of these two areas is maximized.[4]

---

[3]Note that the sum of the two allocations in Figure 5.2 is always 20. The left-hand axis represents an allocation of all 20 units to Period 2, and the right-hand axis represents an allocation entirely to Period 1.

[4]Demonstrate that this point is the maximum by first allocating slightly more to Period 2 (and therefore less to Period 1) and showing that the total area decreases. Conclude by allocating slightly less to Period 2 and showing that, in this case as well, total area declines.

Since we have developed our efficiency criteria independent of an institutional context, these criteria are equally appropriate for evaluating resource allocations generated by markets, government rationing, or even the whims of a dictator. While *any* efficient allocation method must take scarcity into account, the details of precisely how that is done depend on the context.

Intertemporal scarcity imposes an opportunity cost that we henceforth refer to as the *marginal user cost*. When resources are scarce, greater current use diminishes future opportunities. The marginal user cost is the present value of these forgone opportunities at the margin. To be more specific, uses of those resources, which would have been appropriate in the absence of scarcity, may no longer be appropriate once scarcity is present. Using large quantities of water to keep lawns lush and green may be wholly appropriate for an area with sufficiently large replenishable water supplies, but quite inappropriate when it denies drinking water to future generations. Failure to take the higher scarcity value of water into account in the present would lead to inefficiency due to the additional cost resulting from the increased scarcity imposed on the future. This additional marginal value created by scarcity is the marginal user cost.

We can illustrate this concept by returning to our numerical example. With 30 or more units, each period would be allocated 15, the resource would not be scarce, and the marginal user cost would be zero.

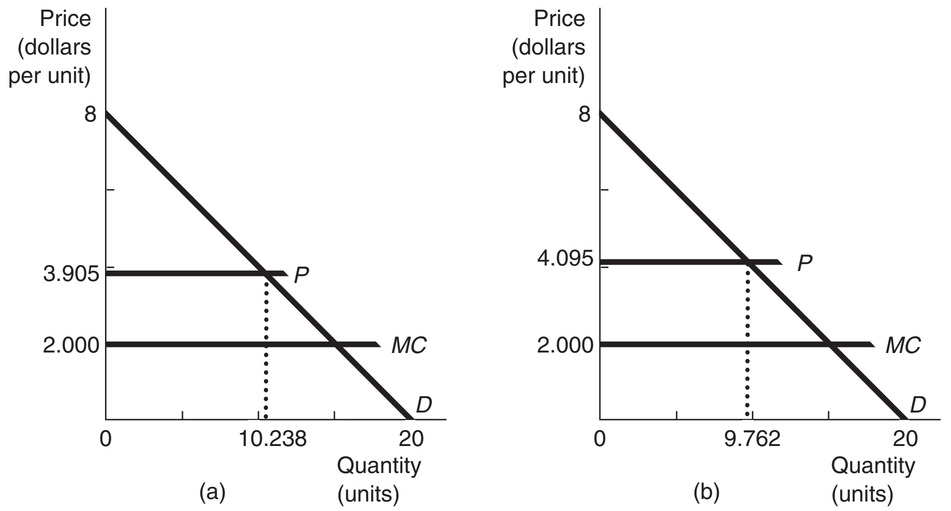
With 20 units, however, scarcity emerges. No longer can 15 units be allocated to each period; each period will have to be allocated less than would be the case without scarcity. Due to this scarcity the marginal user cost for this case is not zero. As can be seen from Figure 5.2, the present value of the marginal user cost, the additional value created by scarcity, is graphically represented by the vertical distance between the quantity axis and the intersection of the two present-value curves. It is identical to the present value of the marginal net benefit in each of the periods. This value can either be read off the graph or determined more precisely, as demonstrated in the chapter appendix, to be \$1.905.

We can make this concept even more concrete by considering its use in a market context. An efficient market would have to consider not only the marginal cost of extraction for this resource but also the marginal user cost. Whereas in the absence of scarcity, the price would equal only the marginal cost of extraction, with scarcity, the price would equal the sum of marginal extraction cost and marginal user cost.

To see this, solve for the prices that would prevail in an efficient market facing scarcity over time. Inserting the efficient quantities (10.238 and 9.762, respectively) into the willingness-to-pay function ($P = 8 - 0.4q$) yields $P_1 = 3.905$ and $P_2 = 4.095$. The corresponding supply-and-demand diagrams are given in Figure 5.3. Compare Figure 5.3 with Figure 5.1 to see the impact of scarcity on price. Note that in the absence of scarcity, marginal user cost is zero.

In an efficient market, the marginal user cost for each period is the difference between the price and the marginal cost of extraction. Notice that it takes the value \$1.905 in the first period and \$2.095 in the second. In both the periods, the present value of the marginal user cost is \$1.905. In the second period, the actual marginal user cost is $\$1.905(1 + r)$. Since $r = 0.10$ in this example, the marginal user cost for

FIGURE 5.3 The Efficient Market Allocation of a Depletable Resource: The Constant-Marginal-Cost Case: (a) Period 1 and (b) Period 2



the second period is \$2.095.[5] Thus, while the present value of marginal user cost is equal in both periods, the actual marginal user cost rises over time.

Both the size of the marginal user cost and the allocation of the resource between the two periods is affected by the discount rate. In Figure 5.2, because of discounting, the efficient allocation allocates somewhat more to Period 1 than to Period 2. A discount rate larger than 0.10 would be incorporated in this diagram by rotating the Period 2 curve an appropriate amount toward the right-hand axis, holding fixed the point at which it intersects the horizontal axis. (Can you see why?) The larger the discount rate, the greater the amount of rotation required. The amount allocated to the second period would be necessarily smaller with larger discount rates. The general conclusion, which holds for all models we consider, is that higher discount rates tend to skew resource extraction toward the present because they give the future less weight in balancing the relative value of present and future resource use. The choice of what discount rate to use, then, becomes a very important consideration for decision makers.

Defining Intertemporal Fairness

While no generally accepted standards of fairness or justice exist, some have more prominent support than others. One such standard concerns the treatment of future generations. What legacy should earlier generations leave to later ones?

[5]You can verify this by taking the present value of \$2.095 and showing it to be equal to \$1.905.

This is a particularly difficult issue because, in contrast to other groups for which we may want to ensure fair treatment, future generations cannot articulate their wishes, much less negotiate with current generations. ("We'll take your radioactive wastes, if you leave us plentiful supplies of titanium.")

One starting point for intergenerational equity is provided by philosopher John Rawls in his monumental work, *A Theory of Justice*. Rawls suggests one way to derive general principles of justice is to place, hypothetically, all people into an original position behind a "veil of ignorance." This veil of ignorance would prevent them from knowing their eventual position in society. Once behind this veil, people would decide on rules to govern the society that they would, after the decision, be forced to inhabit.

In our context, this approach would suggest a hypothetical meeting of all members of present and future generations to decide on rules for allocating resources among generations. Because these members are prevented by the veil of ignorance from knowing the generation to which they will belong, they will not be excessively conservationist (lest they turn out to be a member of an earlier generation) or excessively exploitative (lest they become a member of a later generation).

What kind of rule would emerge from such a meeting? One possibility is the sustainability criterion. The *sustainability criterion* suggests that, at a minimum, future generations should be left no worse off than current generations. Allocations that impoverish future generations, in order to enrich current generations, are, according to this criterion, patently unfair.

In essence, the sustainability criterion suggests that earlier generations are at liberty to use resources that would thereby be denied to future generations as long as the well-being of future generations remains just as high as that of all previous generations. On the other hand, diverting resources from future use would violate the sustainability criterion if it reduced the well-being of future generations below the level enjoyed by preceding generations.

One of the implications of this definition of sustainability is that it is possible for the current generation to use resources (even depletable resources) as long as the interests of future generations could be protected. Do our institutions provide adequate protection for future generations? We begin with examining the conditions under which efficient allocations satisfy the sustainability criterion. Are all efficient allocations sustainable?

Are Efficient Allocations Fair?

In the numerical example we have constructed, it certainly does not appear that the efficient allocation satisfies the sustainable criterion. In the two-period example, more resources are allocated to the first period than to the second. Therefore, net benefits in the second period are lower than in the first. Sustainability does not allow earlier generations to profit at the expense of later generations, and this example certainly appears to be a case where that is happening.

Yet choosing this particular extraction path does not prevent those in the first period from saving some of the net benefits for those in the second period. If the allocation is dynamically efficient, it will always be possible to set aside sufficient net benefits accrued in the first period for those in the second period, so that those in the second period will be at least as well off as they would have been with any other extraction profile.

We can illustrate this point with a numerical example that compares a dynamic efficient allocation with sharing to an allocation where resources are committed equally to each generation. Suppose, for example, you believe that setting aside half (10 units) of the available resources for each period would be a better allocation than the dynamic efficient allocation. The net benefits to each period from this alternative scheme would be $40. Can you see why?

Now let's compare this to an allocation of net benefits that could be achieved with the dynamic efficient allocation. If the dynamic efficient allocation is to satisfy the sustainability criterion, we must be able to show that it can produce an outcome such that each generation would be at least as well off as it would be with the equal allocation. Can that be demonstrated?

In the dynamic efficient allocation, the net benefits to the first period were 40.466, while those for the second period were 39.512.[6] Clearly, if no sharing between the periods took place, this example would violate the sustainability criterion; the second generation is worse off.

But suppose the first generation was willing to share some of the net benefits from the extracted resources with the second generation. If the first generation keeps net benefits of $40 (thereby making it just as well off as if equal amounts were extracted in each period) and saves the extra $0.466 (the $40.466 net benefits earned during the first period in the dynamic efficient allocation minus the $40 reserved for itself) at 10 percent interest for those in the next period, this savings would grow to $0.513 by the second period [0.466(1.10)]. Add this to the net benefits received directly from the dynamic efficient allocation ($39.512), and the second generation would receive $40.025. Those in the second period would be better off by accepting the dynamic efficient allocation with sharing than they would if they demanded that resources be allocated equally between the two periods.

This example demonstrates that although dynamic efficient allocations do not automatically satisfy sustainability criteria, they could be compatible with sustainability, even in an economy relying heavily on depletable resources. The possibility that the second period can be better off is not a guarantee; the required degree of sharing must take place. Example 5.1 points out that this sharing does sometimes take place, although, as we shall see, such sharing is more likely to be the exception rather than the norm. In subsequent chapters, we shall examine both the conditions under which we could expect the appropriate degree of sharing to take place and the conditions under which it would not.

[6]The supporting calculations are (1.905)(10.238) + 0.5(4.095)(10.238) for the first period and (2.095)(9.762) + 0.5(3.905)(9.762) for the second period.

EXAMPLE 5.1

The Alaska Permanent Fund

One interesting example of an intergenerational sharing mechanism currently exists in the State of Alaska. Extraction from Alaska's oil fields generates significant income, but it also depreciates one of the state's main environmental assets. To protect the interests of future generations as the Alaskan pipeline construction neared completion in 1976, Alaska voters approved a constitutional amendment that authorized the establishment of a dedicated fund: the Alaska Permanent Fund. This fund was designed to capture a portion of the rents received from the sale of the state's oil to share with future generations. The amendment requires:

> *At least 25 percent of all mineral lease rentals, royalties, royalty sales proceeds, federal mineral revenue-sharing payments and bonuses received by the state be placed in a permanent fund, the principal of which may only be used for income-producing investments.*

The principal of this fund cannot be used to cover current expenses without a majority vote of Alaskans.

The fund is fully invested in capital markets and diversified among various asset classes. It generates income from interest on bonds, stock dividends, real estate rents, and capital gains from the sale of assets.[7] To date, the legislature has used some of these annual earnings to provide dividends to every eligible Alaska resident, while using the rest to increase the size of the principal, thereby assuring that it is not eroded by inflation. The 2010 dividend was $1,281.

Although this fund does preserve some of the revenue for future generations, two characteristics are worth noting. First, the principal could be used for current expenditures if a majority of current voters agreed. To date, that has not happened, but it has been discussed. Second, only 25 percent of the oil revenue is placed in the fund; assuming that revenue reflects scarcity rent, full sustainability would require dedicating 100 percent of it to the fund. Because the current generation not only gets its share of the income from the permanent fund, but also receives 75 percent of the proceeds from current oil sales, this sharing arrangement falls short of that prescribed by the Hartwick Rule.

Source: The Alaska Permanent Fund Web site: http://www.pfd.state.ak.us/

Applying the Sustainability Criterion

One of the difficulties in assessing the fairness of intertemporal allocations using this version of the sustainability criterion is that it is so difficult to apply. Discovering whether the well-being of future generations is lower than that of current generations requires us not only to know something about the allocation of

---

[7]The fund is managed by the Alaska Permanent Fund Corporation. http://www.apfc.org/home/Content/home/index.cfm

resources over time, but also to know something about the preferences of future generations (in order to establish how valuable various resource streams are to them). That is a tall (impossible?) order!

Is it possible to develop a version of the sustainability criterion that is more operational? Fortunately it is, thanks to what has become known as the "Hartwick Rule." In an early article, John Hartwick (1977) demonstrated that a constant level of consumption could be maintained perpetually from an environmental endowment if all the scarcity rent derived from resources extracted from that endowment were invested in capital. That level of investment would be sufficient to assure that the value of the total capital stock would not decline.

Two important insights flow from this reinterpretation of the sustainability criterion. First, with this version it is possible to judge the sustainability of an allocation by examining whether or not the value of the total capital stock is nondeclining. That test can be performed each year without knowing anything about future allocations or preferences. Second, this analysis suggests the specific degree of sharing that would be necessary to produce a sustainable outcome, namely, all scarcity rent must be invested.

Let's pause to be sure we understand what is being said and why it is being said. Although we shall return to this subject later in the book, it is important now to have at least an intuitive understanding of the implications of this analysis. Consider an analogy. Suppose a grandparent left you an inheritance of $10,000, and you put it in a bank where it earns 10 percent interest.

What are the choices for allocating that money over time and what are the implications of those choices? If you spent exactly $1,000 per year, the amount in the bank would remain $10,000 and the income would last forever; you would be spending only the interest, leaving the principal intact. If you spend more than $1,000 per year, the principal would necessarily decline over time and eventually the balance in the account would go to zero. In the context of this discussion, spending $1,000 per year or less would satisfy the sustainability criterion, while spending more would violate it.

What does the Hartwick Rule mean in this context? It suggests that one way to tell whether an allocation (spending pattern) is sustainable or not is to examine what is happening to the value of the principal over time. If the principal is declining, the allocation (spending pattern) is not sustainable. If the principal is increasing or remaining constant, the allocation (spending pattern) is sustainable.

How do we apply this logic to the environment? In general, the Hartwick Rule suggests that the current generation has been given an endowment. Part of the endowment consists of environmental and natural resources (known as "natural capital") and physical capital (such as buildings, equipment, schools, and roads). Sustainable use of this endowment implies that we should keep the principal (the value of the endowment) intact and live off only the flow of services provided. We should not, in other words, chop down all the trees and use up all the oil, leaving future generations to fend for themselves. Rather we need to assure that the value of the total capital stock is maintained, not depleted.

The desirability of this version of the sustainability criterion depends crucially on how substitutable the two forms of capital are. If physical capital can readily substitute for natural capital, then maintaining the value of the sum of the two is sufficient. If, however, physical capital cannot completely substitute for natural capital, investments in physical capital alone may not be enough to assure sustainability.

How tenable is the assumption of complete substitutability between physical and natural capital? Clearly it is untenable for certain categories of environmental resources. Although we can contemplate the replacement of natural breathable air with universal air-conditioning in domed cities, both the expense and the artificiality of this approach make it an absurd compensation device. Obviously intergenerational compensation must be approached carefully (see Example 5.2).

Recognizing the weakness of the constant total capital definition in the face of limited substitution possibilities has led some economists to propose a new definition. According to this new definition, an allocation is sustainable if it maintains the value of the stock of *natural* capital. This definition assumes that it is natural capital that drives future well-being, and further assumes that little or no substitution between physical and natural capital is possible. To differentiate these

EXAMPLE 5.2

Nauru: Weak Sustainability in the Extreme

The weak sustainability criterion is used to judge whether the depletion of natural capital is offset by sufficiently large increases in physical or financial capital so as to prevent total capital from declining. It seems quite natural to suppose that a violation of that criterion does demonstrate *unsustainable* behavior. But does fulfillment of the weak sustainability criterion provide an adequate test of *sustainable* behavior? Consider the case of Nauru.

Nauru is a small Pacific island that lies some 3,000 kilometers northeast of Australia. It contains one of the highest grades of phosphate rock ever discovered. Phosphate is a prime ingredient in fertilizers.

Over the course of a century, first colonizers and then, after independence, the Nauruans decided to extract massive amounts of this rock. This decision has simultaneously enriched the remaining inhabitants (including the creation of a trust fund believed to contain over $1 billion) and destroyed most of the local ecosystems. Local needs are now mainly met by imports financed from the financial capital created by the sales of the phosphate.

However wise or unwise the choices made by the people of Nauru were, they could not be replicated globally. Everyone cannot subsist solely on imports financed with trust funds; every import must be exported by someone! The story of Nauru demonstrates the value of complementing the weak sustainability criterion with other, more demanding criteria. Satisfying the weak sustainability criterion may be a necessary condition for sustainability, but it is not always sufficient.

Source: J. W. Gowdy and C. N. McDaniel, "The Physical Destruction of Nauru: An Example of Weak Sustainability." LAND ECONOMICS, Vol. 75, No. 2 (1999), pp. 333–338.

two definitions, the maintenance of the value of total capital is known as the "weak sustainability" definition, while maintaining the value of natural capital is known as the "strong sustainability" definition.

A final definition, known as "environmental sustainability," requires that certain *physical flows* of certain key *individual* resources be maintained. This definition suggests that it is not sufficient to maintain the *value* of an *aggregate*. For a fishery, for example, this definition would require catch levels that did not exceed the growth of the biomass for the fishery. For a wetland, it would require the preservation of the specific ecological functions.

Implications for Environmental Policy

In order to be useful guides to policy, our sustainability and efficiency criteria must be neither synonymous nor incompatible. Do these criteria meet that test?

They do. As we shall see later in the book, not all efficient allocations are sustainable and not all sustainable allocations are efficient. Yet some sustainable allocations are efficient and some efficient allocations are sustainable. Furthermore, market allocations may be either efficient or inefficient and either sustainable or unsustainable.

Do these differences have any policy implications? Indeed they do. In particular they suggest a specific strategy for policy. Among the possible uses for resources that fulfill the sustainability criterion, we choose the one that maximizes either dynamic or static efficiency as appropriate. In this formulation the sustainability criterion acts as an overriding constraint on social decisions. Yet by itself, the sustainability criterion is insufficient because it fails to provide any guidance on which of the infinite number of sustainable allocations should be chosen. That is where efficiency comes in. It provides a means for maximizing the wealth derived from all the possible sustainable allocations.

This combination of efficiency with sustainability turns out to be very helpful in guiding policy. Many unsustainable allocations are the result of inefficient behavior. Correcting the inefficiency can either restore sustainability or move the economy a long way in that direction. Furthermore, and this is important, correcting inefficiencies can frequently produce win-win situations. In win-win changes, the various parties affected by the change can all be made better off after the change than before. This contrasts sharply with changes in which the gains to the gainers are smaller than the losses to the losers.

Win-win situations are possible because moving from an inefficient to an efficient allocation increases net benefits. The increase in net benefits provides a means for compensating those who might otherwise lose from the change. Compensating losers reduces the opposition to change, thereby making change more likely. Do our economic and political institutions normally produce outcomes that are both efficient and sustainable? In upcoming chapters we will provide explicit answers to this important question.

Summary

Both efficiency and ethical considerations can guide the desirability of private and social choices involving the environment. Whereas the former is concerned mainly with eliminating waste in the use of resources, the latter is concerned with assuring the fair treatment of all parties.

Previous chapters have focused on the static and dynamic efficiency criteria. Chapter 20 will focus on the environmental justice implications of environmental degradation and remediation for members of the current generation. This chapter examines one globally important characterization of the obligation previous generations owe to those generations that follow and the policy implications that flow from acceptance of that obligation.

The specific obligation examined in this chapter—sustainable development—is based upon the notion that earlier generations should be free to pursue their own well-being as long as in so doing they do not diminish the welfare of future generations. This notion gives rise to three alternative definitions of sustainable allocations:

Weak Sustainability. Resource use by previous generations should not exceed a level that would prevent subsequent generations from achieving a level of well-being at least as great. One of the implications of this definition is that the value of the capital stock (natural plus physical capital) should not decline. Individual components of the aggregate could decline in value as long as other components were increased in value (normally through investment) sufficiently to leave the aggregate value unchanged.

Strong Sustainability. According to this interpretation, the value of the remaining stock of natural capital should not decrease. This definition places special emphasis on preserving natural (as opposed to total) capital under the assumption that natural and physical capital offer limited substitution possibilities. This definition retains the focus of the previous definition on preserving value (rather than a specific level of physical flow) and on preserving an aggregate of natural capital (rather than any specific component).

Environmental Sustainability. Under this definition, the *physical* flows of *individual* resources should be maintained, not merely the *value* of the *aggregate*. For a fishery, for example, this definition would emphasize maintaining a constant fish catch (referred to as a sustainable yield), rather than a constant value of the fishery. For a wetland, it would involve preserving specific ecological functions, not merely their aggregate value.

It is possible to examine and compare the theoretical conditions that characterize various allocations (including market allocations and efficient allocations) to the necessary conditions for an allocation to be sustainable under these definitions. According to the theorem that is now known as the "Hartwick Rule," if all of the scarcity rent from the use of scarce resources is invested in capital, the resulting allocation will satisfy the first definition of sustainability.

In general, not all efficient allocations are sustainable and not all sustainable allocations efficient. Furthermore, market allocations can be: (1) efficient, but not sustainable; (2) sustainable, but not efficient; (3) inefficient and unsustainable; and (4) efficient and sustainable. One class of situations, known as “win-win” situations, provides an opportunity to increase simultaneously the welfare of both current and future generations.

We shall explore these themes much more intensively as we proceed through the book. In particular we shall inquire into when market allocations can be expected to produce allocations that satisfy the sustainability definitions and when they cannot. We shall also see how the skillful use of economic incentives can allow policy-makers to exploit “win-win” situations to promote a transition onto a sustainable path for the future.

Discussion Question

1. The environmental sustainability criterion differs in important ways from both strong and weak sustainability. Environmental sustainability frequently means maintaining a constant physical flow of individual resources (e.g., fish from the sea or wood from the forest), while the other two definitions call for maintaining the *aggregate value* of those service flows. When might the two criteria lead to different choices? Why?

Self-Test Exercises

1. In the numerical example given in the text, the inverse demand function for the depletable resource is $P = 8 - 0.4q$ and the marginal cost of supplying it is \$2. (a) If 20 units are to be allocated between two periods, in a dynamic efficient allocation how much would be allocated to the first period and how much to the second period when the discount rate is zero? (b) Given this discount rate what would be the efficient price in the two periods? (c) What would be the marginal user cost in each period?
2. Assume the same demand conditions as stated in Problem 1, but let the discount rate be 0.10 and the marginal cost of extraction be \$4. How much would be produced in each period in an efficient allocation? What would be the marginal user cost in each period? Would the static and dynamic efficiency criteria yield the same answers for this problem? Why?
3. Compare two versions of the two-period depletable resource model that differ only in the treatment of marginal extraction cost. Assume that in the second version the constant marginal extraction cost is lower in the second period than the first (perhaps due to the anticipated arrival of a new, superior extraction technology). The constant marginal extraction cost is the same in both periods in the first version and is equal to the marginal extraction cost in the first period of the second version. In a dynamic efficient allocation, how would the extraction profile in the second version differ from the first?

Would relatively more or less be allocated to the second period in the second version than in the first version? Would the marginal user cost be higher or lower in the second version? Why?

4. a. Consider the general effect of the discount rate on the dynamic efficient allocation of a depletable resource across time. Suppose we have two versions of the two-period model discussed in this chapter. The two versions are identical except for the fact that the second version involves a higher discount rate than the first version. What effect would the higher discount rate have on the allocation between the two periods and the magnitude of the present value of the marginal user cost?
   b. Explain the intuition behind your results.
5. a. Consider the effect of population growth on the allocation on the dynamic efficient allocation of a depletable resource across time. Suppose we have two versions of the two-period model, discussed in this chapter, that are identical except for the fact that the second version involves a higher demand for the resource in the second period (the demand curve shifts to the right due to population growth) than the first version. What effect would the higher demand in the second period have on the allocation between the two periods and the magnitude of the present value of the marginal user cost?
   b. Explain the intuition behind your results.

Further Reading

Atkinson, G. et al. *Measuring Sustainable Development: Macroeconomics and the Environment* (Cheltenham, UK: Edward Elgar, 1997). Tackles the tricky question of how one can tell whether development is sustainable or not.

Desimone, L. D. *Eco-Efficiency: The Business Link to Sustainable Development* (Cambridge, MA: MIT Press, 1997). What is the role for the private sector in sustainable development? Is concern over the "bottom line" consistent with the desire to promote sustainable development?

May, P., and R.S.D. Motta, eds. *Pricing the Planet: Economic Analysis for Sustainable Development* (New York: Columbia University Press, 1996). Ten essays dealing with how sustainable development might be implemented.

Perrings, C. *Sustainable Development and Poverty Alleviation in Sub-Saharan Africa: The Case of Botswana* (New York: Macmillan, 1996). A leading practitioner in the field examines the problems and prospects for sustainable development in Botswana.

Pezzey, J.V.C., and Michael A. Toman. "Progress and Problems in the Economics of Sustainability," in T. Tietenberg and H. Folmer, eds. *The International Yearbook of Environmental and Resource Economics: A Survey of Current Issues* (Cheltenham, UK: Edward Elgar, 2002): 165–232. An excellent survey of the economics literature on sustainable development.

Additional References and Historically Significant References are available on this book's Companion Website: http://www.pearsonhighered.com/tietenberg/

Appendix

The Mathematics of the Two-Period Model

An exact solution to the two-period model can be derived using the solution equations derived in the appendix to Chapter 3.

The following parameter values are assumed by the two-period example:

$$a = 8,\ c = \$2,\ b = 0.4, Q = 20, \text{ and } r = 0.10.$$

Using these, we obtain

$$8 - 0.4q_1 - 2 - \lambda = 0, \tag{1}$$

$$\frac{8 - 0.4q_2 - 2}{1.10} - \lambda = 0 \tag{2}$$

$$q_1 + q_2 = 20.$$

It is now readily verified that the solution (accurate to the third decimal place) is

$$q_1 = 10.238,\ q_2 = 9.762,\ \lambda = \$1.905.$$

We can now demonstrate the propositions discussed in this text.

1. Verbally, Equation (1) states that in a dynamic efficient allocation, the present value of the marginal net benefit in Period 1 ($8 - 0.4q_1 - 2$) has to equal λ. Equation (2) states that the present value of the marginal net benefit in Period 2 should also equal λ. Therefore, they must equal each other. This demonstrates the proposition shown graphically in Figure 5.2.
2. The present value of marginal user cost is represented by λ. Thus, Equation (1) states that price in the first period ($8 - 0.4q_1$) should be equal to the sum of marginal extraction cost ($2) and marginal user cost ($1.905). Multiplying (2) by $1 + r$, it becomes clear that price in the second period ($8 - 0.4q_2$) is equal to the marginal extraction cost ($2) plus the higher marginal user cost [$\lambda(1 + r) = (1.905)(1.10) = \$2.095$] in Period 2. These results show why the graphs in Figure 5.3 have the properties they do. They also illustrate the point that, in this case, marginal user cost rises over time.

6 Depletable Resource Allocation: The Role of Longer Time Horizons, Substitutes, and Extraction Cost

The whole machinery of our intelligence, our general ideas and laws, fixed and external objects, principles, persons, and gods, are so many symbolic, algebraic expressions. They stand for experience; experience which we are incapable of retaining and surveying in its multitudinous immediacy. We should flounder hopelessly, like the animals, did we not keep ourselves afloat and direct our course by these intellectual devices. Theory helps us to bear our ignorance of fact.

—George Santayana, *The Sense of Beauty* (1896)

Introduction

How do societies react when finite stocks of depletable resources become scarce? Is it reasonable to expect that self-limiting feedbacks would facilitate the transition to a sustainable, steady state? Or is it more reasonable to expect that self-reinforcing feedback mechanisms would cause the system to overshoot the resource base, possibly even precipitating a societal collapse?

We begin to seek answers to these questions by studying the implications of both efficient and profit-maximizing decision making. What kinds of feedback mechanisms are implied by decisions motivated by efficiency and by profit maximization? Are they compatible with a smooth transition or are they more likely to produce overshoot and collapse?

We approach these questions in several steps, first by defining and discussing a simple but useful *resource taxonomy* (classification system), as well as explaining the dangers of ignoring the distinctions made by this taxonomy. We begin the analysis by defining an efficient allocation of an exhaustible resource over time when no renewable substitute is available. This is accomplished by exploring the conditions any efficient allocation must satisfy and using numerical examples to illustrate the implications of these conditions.

Renewable resources are integrated into the analysis by relying on the simplest possible case—the resource is supplied at a fixed, abundant rate and can be accessed at a constant marginal cost. Solar energy and replenishable surface water are two examples that seem roughly to fit this characterization. Combining this model of renewable resource supply with our basic depletable resource model allows us to characterize efficient extraction paths for both types of resources, assuming that they are perfect substitutes. We explore how these efficient paths are affected by changes in the nature of the cost functions as well as the nature of market extraction paths. Whether or not the market is capable of yielding a dynamically efficient allocation in the presence or absence of a renewable substitute provides a focal point for the analysis. Succeeding chapters will use these principles to examine the allocation of energy, food, and water resources, and to provide a basis for developing more elaborate models of renewable biological populations, such as fisheries and forests.

A Resource Taxonomy

Three separate concepts are used to classify the stock of depletable resources: (1) *current reserves*, (2) *potential reserves*, and (3) *resource endowment*. The U.S. Geological Survey (USGS) has the official responsibility for keeping records of the U.S. resource base, and has developed the classification system described in Figure 6.1.

Note the two dimensions—one economic and one geological. A movement from top to bottom represents movement from cheaply extractable resources to those extracted at substantially higher costs. By contrast, a movement from left to right represents increasing geological uncertainty about the size of the resource base.

Current reserves (shaded area in Figure 6.1) are defined as known resources that can profitably be extracted at current prices. The magnitude of these current reserves can be expressed as a number.

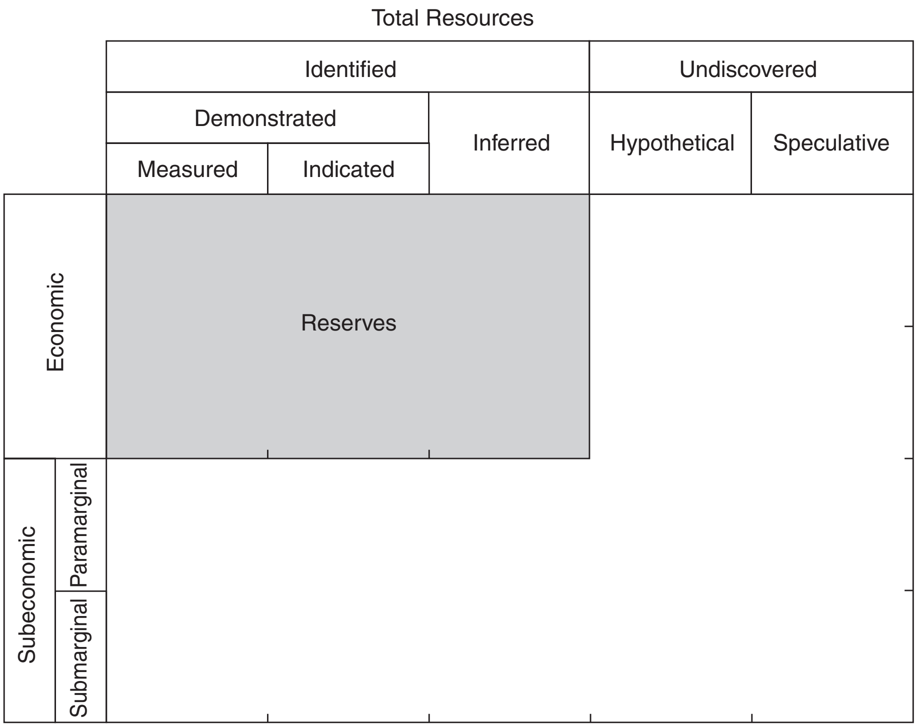
Potential reserves, on the other hand, are most accurately defined as a function rather than a number. The amount of reserves potentially available depends upon the price people are willing to pay for those resources—the higher the price, the larger the potential reserves. For example, studies examining the amount of additional oil that could be recovered from existing oil fields by enhanced recovery techniques, such as injecting solvents or steam into the well to lower the density of the oil, find that as price is increased, the amount of economically recoverable oil also increases.

The *resource endowment* represents the natural occurrence of resources in the earth's crust. Since prices have nothing to do with the size of the resource endowment, it is a geological, rather than an economic, concept. This concept is important because it represents an upper limit on the availability of terrestrial resources.

The distinctions among these three concepts are significant. One common mistake in failing to respect these distinctions is using data on current reserves as if they represented the maximum potential reserves. This fundamental error can cause a huge understatement of the time until exhaustion.

A second common mistake is to assume that the entire resource endowment can be made available as potential reserves at a price people would be willing to pay. Clearly, if an infinite price were possible, the entire resource endowment could be exploited. However, an infinite price is not likely.

FIGURE 6.1 A Categorization of Resources



Terms

Identified resources: specific bodies of mineral-bearing material whose location, quality, and quantity are known from geological evidence, supported by engineering measurements.

Measured resources: material for which quantity and quality estimates are within a margin of error of less than 20 percent, from geologically well-known sample sites.

Indicated resources: material for which quantity and quality have been estimated partly from sample analyses and partly from reasonable geological projections.

Inferred resources: material in unexplored extensions of demonstrated resources based on geological projections.

Undiscovered resources: unspecified bodies of mineral-bearing material surmised to exist on the basis of broad geological knowledge and theory.

Hypothetical resources: undiscovered materials reasonably expected to exist in a known mining district under known geological conditions.

Speculative resources: undiscovered materials that may occur in either known types of deposits in favorable geological settings where no discoveries have been made, or in yet unknown types of deposits that remain to be recognized.

Source: U.S. Bureau of Mines and the U.S. Geological Survey, "Principle of the Mineral Resource Classification System of the U.S. Bureau of Mines and the U.S. Geological Survey." GEOLOGICAL SURVEY BULLETIN, 1976, pp. 1450-A.

Other distinctions among resource categories are also useful. The first category includes all depletable, recyclable resources, such as copper. A *depletable resource* is one for which the natural replenishment feedback loop can safely be ignored. The rate of replenishment for these resources is so low that it does not offer a potential for augmenting the stock in any reasonable time frame.

A *recyclable resource* is one that, although currently being used for some particular purpose, exists in a form allowing its mass to be recovered once that purpose is no longer necessary or desirable. For example, copper wiring from an automobile can be recovered after the car has been shipped to the junkyard. The degree to which a resource is recycled is determined by economic conditions, a subject covered in Chapter 8.

The current reserves of a depletable, recyclable resource can be augmented by economic replenishment, as well as by recycling. Economic replenishment takes many forms, all sharing the characteristic that they turn previously unrecoverable resources into recoverable ones. One obvious stimulant for this replenishment is price. As price rises, producers find it profitable to explore more widely, dig more deeply, and use lower-concentration ores.

Higher prices also stimulate technological progress. Technological progress simply refers to an advancement in the state of knowledge that allows us to expand the set of feasible possibilities. One profound, if controversial, example can be found in the successful harnessing of nuclear power.

The other side of the coin for depletable, recyclable resources is that their potential reserves can be exhausted. The depletion rate is affected by the demand for and the durability of the products built with the resource, and the ability to reuse the products. Except where demand is totally price-inelastic (i.e., insensitive to price), higher prices tend to reduce the quantity demanded. Durable products last longer, reducing the need for newer ones. Reusable products provide a substitute for new products. In the commercial sector, reusable soft drink containers provide one example, while flea markets (where secondhand items are sold) provide another for the household sector.

For some resources, the size of the potential reserves depends explicitly on our ability to store the resource. For example, helium is generally found commingled with natural gas in common fields. As the natural gas is extracted and stored, unless the helium is simultaneously captured and stored, it diffuses into the atmosphere. This diffusion results in such low concentrations that extraction of helium from the air is not economical at current or even likely future prices. Thus, the useful stock of helium depends crucially on how much we decide to store.

Not all depletable resources can be recycled or reused. Depletable energy resources such as coal, oil, and gas are consumed as they are used. Once combusted and turned into heat energy, the heat dissipates into the atmosphere and becomes nonrecoverable.

The endowment of depletable resources is of finite size. Current use of depletable, nonrecyclable resources precludes future use; hence, the issue of how they should be shared among generations is raised in the starkest, least forgiving form.

Depletable, recyclable resources raise this same issue, though somewhat less starkly. Recycling and reuse make the useful stock last longer, all other things being equal. It is tempting to suggest that depletable, recyclable resources could last forever with 100 percent recycling, but unfortunately the physical theoretical upper limit on recycling is less than 100 percent—an implication of a version of the entropy law defined in Chapter 2. Some of the mass is always lost during recycling or use.

As long as less than 100 percent of the mass is recycled, the useful stock must eventually decline to zero. Even for recyclable, depletable resources, the cumulative useful stock is finite, and current consumption patterns still have an effect on future generations.

Renewable resources are differentiated from depletable resources primarily by the fact that natural replenishment augments the flow of renewable resources at a nonnegligible rate. Solar energy, water, cereal grains, fish, forests, and animals are all examples of renewable resources. Thus, it is possible, though not inexorable, that a flow of these resources could be maintained perpetually.[1]

For some renewable resources, the continuation and volume of their flow depend crucially on humans. Soil erosion and nutrient depletion reduce the flow of food. Excessive fishing reduces the stock of fish, which in turn reduces the rate of natural increase of the fish population. Other examples abound.

For other renewable resources, such as solar energy, the flow is independent of humans. The amount consumed by one generation does not reduce the amount that can be consumed by subsequent generations.

Some renewable resources can be stored; others cannot. For those that can, storage provides a valuable way to manage the allocation of the resource over time. We are not left simply at the mercy of natural ebbs and flows. Food without proper care perishes rapidly, but under the right conditions storage can be used to feed the hungry in times of famine. Unstored solar energy radiates off the earth's surface and dissipates into the atmosphere. While solar energy can be stored in many forms, the most common natural form of storage occurs when it is converted to biomass by photosynthesis.

Storage of renewable resources usually performs a different service from storage of depletable resources. Storing depletable resources extends their economic life; storing renewable resources, on the other hand, can serve as a means of smoothing out the cyclical imbalances of supply and demand. Surpluses are stored for use during periods when deficits occur. Food stockpiles and the use of dams to store water to be used for hydropower are two familiar examples.

Managing renewable resources presents a different challenge from managing depletable resources, although an equally significant one. The challenge for depletable resources involves allocating dwindling stocks among generations while meeting the ultimate transition to renewable resources. In contrast, the challenge for managing renewable resources involves the maintenance of an efficient, sustainable flow. Chapters 7 through 13 deal with how the economic and political sectors have responded to these challenges for particularly significant types of resources.

---

[1]Even renewable resources are ultimately finite because their renewability depends on energy from the sun and the sun is expected to serve as an energy source for only the next five or six billion years. That fact does not eliminate the need to manage resources effectively until that time. Furthermore, the finiteness of renewable resources is sufficiently far into the future to make the distinction useful.

Efficient Intertemporal Allocations

If we are to judge the adequacy of market allocations, we must define what is meant by efficiency in relation to the management of depletable and renewable resource allocations. Because allocation over time is the crucial issue, dynamic efficiency becomes the core concept. The dynamic efficiency criterion assumes that society's objective is to maximize the present value of net benefits coming from the resource. For a depletable, nonrecyclable resource, this requires a balancing of the current and subsequent uses of the resource. In order to recall how the dynamic efficiency criterion defines this balance, we shall begin with an elaboration of the very simple two-period model developed in Chapter 5. We shall show how these conclusions generalize to longer planning horizons and more complicated situations.

The Two-Period Model Revisited

In Chapter 5 we defined a situation involving the allocation over two periods of a finite resource that could be extracted at constant marginal cost. With a stable demand curve for the resource, an efficient allocation meant that more than half of the resource was allocated to the first period and less than half to the second period. This allocation was affected both by the marginal cost of extraction and by the marginal user cost.

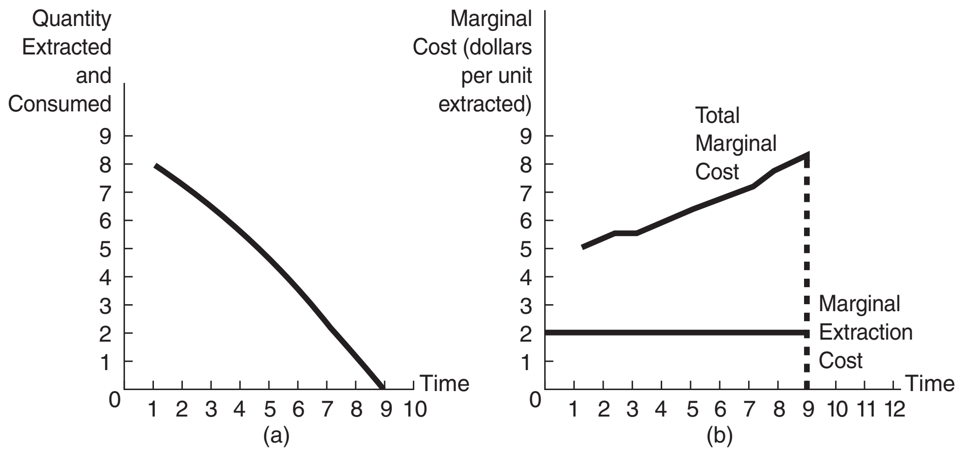
Due to the fixed and finite supplies of depletable resources, production of a unit today precludes production of that unit tomorrow. Therefore, production decisions today must take forgone future net benefits into account. Marginal user cost is the opportunity cost measure that allows intertemporal balancing to take place.

In the two-period model, the marginal cost of extraction is assumed to be constant, but the value of the marginal user cost rises over time. In fact, as was demonstrated mathematically in the appendix to Chapter 5, when the demand curve is stable over time and the marginal cost of extraction is constant, the rate of increase in the current value of the marginal user cost is equal to r, the discount rate. Thus, in Period 2, the marginal user cost would be $1 + r$ times as large as it was in Period 1.[2] Marginal user cost rises at rate r in an efficient allocation in order to preserve the balance between present versus future production.

In summary, our two-period example suggests that an efficient allocation of a finite resource with a constant marginal cost of extraction involves rising marginal user cost and falling quantities consumed. We can now generalize to longer time periods and different extraction circumstances.

[2]The condition that marginal user cost rises at rate r is true only when the marginal cost of extraction is constant. Later in this chapter we show how the marginal user cost is affected when marginal extraction cost is not constant.

FIGURE 6.2 (a) Constant Marginal Extraction Cost with No Substitute Resource: Quantity Profile (b) Constant Marginal Extraction Cost with No Substitute Resource: Marginal Cost Profile



The *N*-Period Constant-Cost Case

We begin this generalization by retaining the constant-marginal-extraction-cost assumption while extending the time horizon within which the resource is allocated. In the numerical example shown in Figures 6.2a and 6.2b, the demand curves and the marginal cost curve from the two-period case are retained. The only changes in this numerical example from the two-period case involve spreading the allocation over a larger number of years and increasing the total recoverable supply from 20 to 40. (The mathematics behind this and subsequent examples is presented in the appendix at the end of this chapter.)

Figure 6.2a demonstrates how the efficient quantity extracted varies over time, while Figure 6.2b shows the behavior of the marginal user cost and the marginal cost of extraction. Total marginal cost refers to the sum of the two. The marginal cost of extraction is represented by the lower line, and the marginal user cost is depicted as the vertical distance between the marginal cost of extraction and the total marginal cost. To avoid confusion, note that the horizontal axis is defined in terms of time, not the more conventional designation—quantity.

Several trends are worth noting. First of all, in this case, as in the two-period case, the efficient marginal user cost rises steadily in spite of the fact that the marginal cost of extraction remains constant. This rise in the efficient marginal user cost reflects increasing scarcity and the accompanying rise in the opportunity cost of current consumption as the remaining stock dwindles.

In response to these rising costs over time, the extracted quantity falls over time until it finally goes to zero, which occurs precisely at the moment when the total marginal cost becomes $8. At this point, total marginal cost is equal to the highest price anyone is willing to pay, so demand and supply simultaneously equal zero.

Thus, even in this challenging case involving no increase in the cost of extraction, an efficient allocation envisions a smooth transition to the exhaustion of a resource. The resource does not "suddenly" run out, although in this case it does run out.

Transition to a Renewable Substitute

So far we have discussed the allocation of a depletable resource when no substitute is available to take its place. Suppose, however, we consider the nature of an efficient allocation when a substitute renewable resource is available at constant marginal cost. This case, for example, could describe the efficient allocation of oil or natural gas with a solar substitute or the efficient allocation of exhaustible groundwater with a surface-water substitute. How could we define an efficient allocation in this circumstance?

Since this problem is very similar to the one already discussed, we can use what we have already learned as a foundation for mastering this new situation. The depletable resource would be exhausted in this case, just as it was in the previous case, but that will be less of a problem, since we'll merely switch to the renewable one at the appropriate time. For the purpose of our numerical example, assume the existence of a perfect substitute for the depletable resource that is infinitely available at a cost of $6 per unit. The transition from the depletable resource to this renewable resource would ultimately transpire because the renewable resource marginal cost ($6) is less than the maximum willingness to pay ($8). (Can you figure out what the efficient allocation would be if the marginal cost of this substitute renewable resource was $9, instead of $6?)

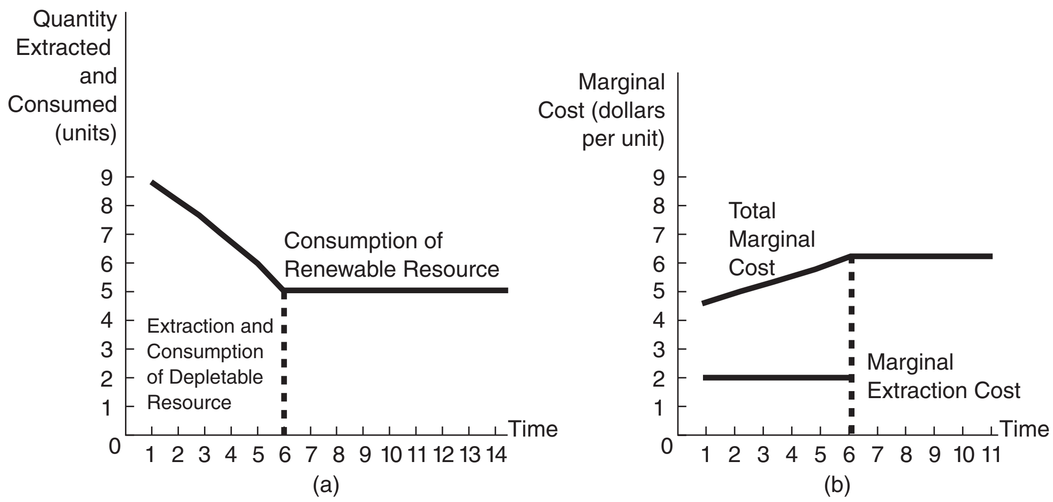
The total marginal cost for the depletable resource in the presence of a $6 perfect substitute would never exceed $6, because society could always use the renewable resource instead, whenever it was cheaper. Thus, while the maximum willingness to pay (the *choke price*) sets the upper limit on total marginal cost when no substitute is available, the marginal cost of extraction of the substitute sets the upper limit when a perfect substitute is available at a marginal cost lower than the choke price. The efficient path for this situation is given in Figures 6.3a and 6.3b.

In this efficient allocation, the transition is once again smooth. Quantity extracted per unit of time is gradually reduced as the marginal user cost rises until the switch is made to the substitute. No abrupt change is evident in either marginal cost or quantity profiles.

Because the renewable resource is available, more of the depletable resource would be extracted in the earlier periods than was the case in our previous numerical example without a renewable resource. As a result, the depletable resource would be exhausted sooner than it would have been without the renewable resource substitute. In this example, the switch is made during the sixth period, whereas in the last example the last units were exhausted at the end of the eighth period. That seems consistent with common sense. When a substitute is available, the need to save some of the depletable resource for the future is certainly less pressing (in other words, the opportunity cost has fallen).

At the transition point, called the switch point, consumption of the renewable resource begins. Prior to the switch point, only the depletable resource is

FIGURE 6.3 (a) Constant Marginal Extraction Cost with Substitute Resource: Quantity Profile (b) Constant Marginal Extraction Cost with Substitute Resource: Marginal Cost Profile

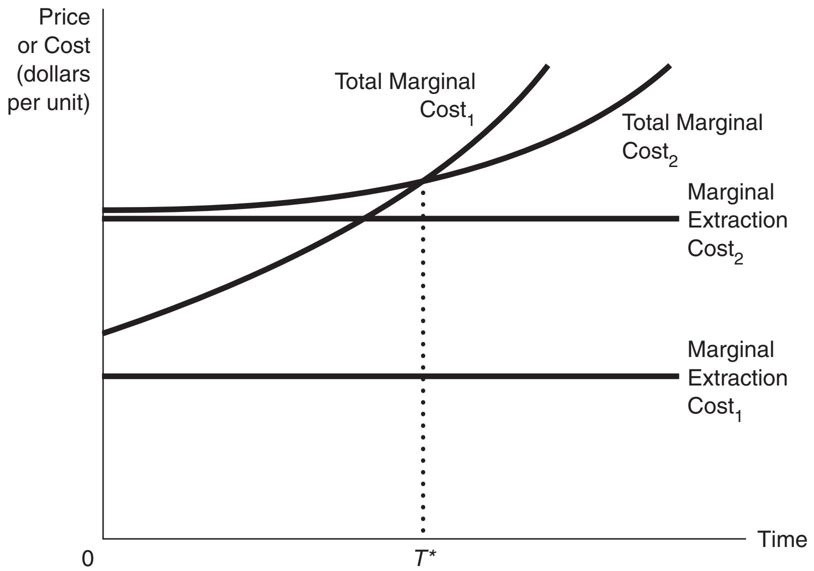


consumed, while after the switch point only the renewable resource is consumed. This sequencing of consumption pattern results from the cost patterns. Prior to the switch point, the depletable resource is cheaper. At the switch point, the marginal cost of the depletable resource (including marginal user cost) rises to meet the marginal cost of the substitute, and the transition occurs. Due to the availability of the substitute resource, after the switch point consumption never drops below five units in any time period. This level is maintained because five is the amount that maximizes the net benefit when the marginal cost equals \$6 (the price of the substitute). (Convince yourself of the validity of this statement by substituting \$6 into the willingness-to-pay function and solving for the quantity demanded.)

We shall not show the numerical example here, but it is not difficult to see how an efficient allocation would be defined when the transition is from one constant marginal-cost depletable resource to another depletable resource with a constant, but higher, marginal cost (see Figure 6.4). The total marginal cost of the first resource would rise over time until it equaled that of the second resource at the time of transition (T^*). In the period of time prior to transition, only the cheapest resource would be consumed; all of it would have been consumed by T^*.

A close examination of the total-marginal-cost path reveals two interesting characteristics worthy of our attention. First, even in this case, the transition is a smooth one; total marginal cost never jumps to the higher level. Second, the slope of the total marginal cost curve over time is flatter after the time of transition.

FIGURE 6.4 The Transition from One Constant-Cost Depletable Resource to Another



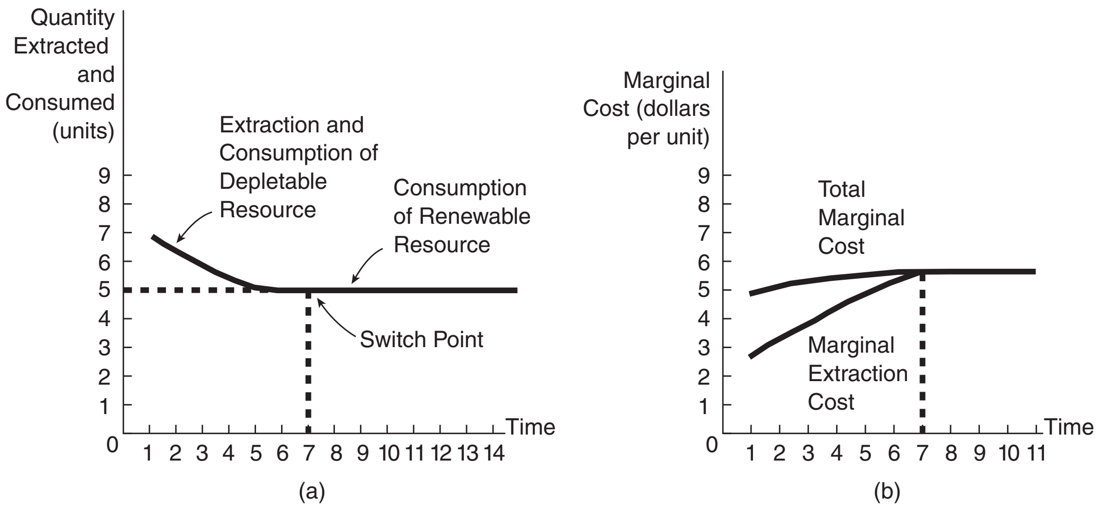
The first characteristic is easy to explain. The total marginal costs of the two resources have to be equal at the time of transition. If they weren't equal, the net benefit could be increased by switching to the lower-cost resource from the more expensive resource. Total marginal costs are not equal in the other periods. In the period before transition, the first resource is cheaper and therefore used exclusively, whereas after transition the first resource is exhausted, leaving only the second resource.

The slope of the marginal cost curve over time is flatter after transition simply because the component of total marginal cost that is growing (the marginal user cost) represents a smaller portion of the total marginal cost of the second resource than of the first. The total marginal cost of each resource is determined by the marginal extraction cost plus the marginal user cost. In both cases the marginal user cost is increasing at rate r, and the marginal cost of extraction is constant. As seen in Figure 6.4, the marginal cost of extraction, which is constant, constitutes a much larger proportion of total marginal cost for the second resource than for the first. Hence, total marginal cost rises more slowly for the second resource, at least initially.

Increasing Marginal Extraction Cost

We have now expanded our examination of the efficient allocation of depletable resources to include longer time horizons and the availability of other depletable or renewable resources that could serve as perfect substitutes. As part of our trek toward increasing realism, we will consider a situation in which the marginal cost of extracting the depletable resource rises with the cumulative amount extracted.

FIGURE 6.5 (a) Increasing Marginal Extraction Cost with Substitute Resource: Quantity Profile
(b) Increasing Marginal Extraction Cost with Substitute Resource: Marginal Cost Profile



This is commonly the case, for example, with minerals, where the higher-grade ores are extracted first, followed by an increasing reliance on lower-grade ones.

Analytically, this case is handled in the same manner as the previous case, except that the function describing the marginal cost of extraction is slightly more complicated.[3] It increases with the cumulative amount extracted. The dynamic efficient allocation of this resource is found by maximizing the present value of the net benefits, using this modified cost of extraction function. The results of that maximization are portrayed in Figures 6.5a and 6.5b.

The most significant difference between this case and the others lies in the behavior of marginal user cost. In the previous case we noted that marginal user cost rose over time at rate r. When the marginal cost of extraction increases with the cumulative amount extracted, marginal user cost *declines* over time until, at the time of transition to the renewable resource, it goes to zero. Why?

Remember that marginal user cost is an opportunity cost reflecting forgone future marginal net benefits. In contrast to the constant marginal-cost case, in the increasing marginal-cost case every unit extracted now raises the cost of future extraction. Therefore, as the current marginal cost rises over time, the sacrifice made by future generations (as an additional unit is consumed earlier) diminishes; the net benefit that would be received by a future generation, if a unit of the resource were saved for them, gets smaller and smaller as the marginal extraction cost of that resource gets larger and larger. By the last period, the marginal extrac-

[3]The marginal cost of extraction is $MC_t = \$2 + 0.1Q_t$ where Q_t is cumulative extraction to date.

tion cost is so high that earlier consumption of one more unit imposes virtually no sacrifice at all. At the switch point, the opportunity cost of current extraction drops to zero, and total marginal cost equals the marginal extraction cost.[4]

The increasing-cost case differs from the constant-cost case in another important way as well. In the constant-cost case, the depletable resource reserve is ultimately completely exhausted. In the increasing-cost case, however, the reserve is not exhausted; some is left in the ground because it is more expensive to use than the substitute.

Up to this point in our analysis, we have examined how an efficient allocation would be defined in a number of circumstances. First, we examined a situation in which a finite amount of a resource is extracted at constant marginal cost. Despite the absence of increasing extraction cost, an efficient allocation involves a smooth transition to a substitute, when one is available, or to abstinence, when one is not. The complication of increasing marginal cost changes the time profile of the marginal user cost, but it does not alter the basic finding of declining consumption of depletable resources coupled with rising total marginal cost.

Can this analysis be used as a basis for judging whether current extraction profiles are efficient? As a look at the historical record reveals, the consumption patterns of most depletable resources have involved increases, not decreases, in consumption over time. Is this prima facie evidence that the resources are not being allocated efficiently?

Exploration and Technological Progress

Using the historical patterns of increasing consumption to conclude that depletable resources are not being allocated efficiently would not represent a valid finding. The models considered to this point have not yet included a consideration of the role of population and income growth, or of the exploration for new resources or technological progress—historically significant factors in the determination of actual consumption paths.

The search for new resources is expensive. As easily discovered resources are exhausted, searches are initiated in less rewarding environments, such as the bottom of the ocean or locations deep within the earth. This suggests the *marginal cost of exploration*, which is the marginal cost of finding additional units of the resource, should be expected to rise over time, just as the marginal cost of extraction does.

As the total marginal cost for a resource rises over time, society should actively explore possible new sources of that resource. Larger increases in the marginal cost of extraction for known sources trigger larger potential increases in net benefits from finding new sources that previously would have been unprofitable to extract.

---

[4]Total marginal cost cannot be greater than the marginal cost of the substitute. Yet, in the increasing marginal extraction cost case, at the time of transition the marginal extraction cost also must equal the marginal cost of the substitute. If that weren't true, it would imply that some of the resource that was available at a marginal cost lower than the substitute would not be used. This would clearly be inefficient, since net benefits could be increased by simply using less of the more expensive substitute. Hence, at the switch point, in the rising marginal-cost case, the marginal extraction cost has to equal total marginal cost, implying a zero marginal user cost.

Some of this exploration would be successful; new sources of the resource would be discovered. If the marginal extraction cost of the newly discovered resources is low enough, these discoveries could lower, or at least delay, the increase in the total marginal cost of production. As a result, the new finds would tend to encourage more consumption. Compared to a situation with no exploration possible, the model with exploration would show a smaller and slower decline in consumption, while the rise in total marginal cost would be dampened.

It is also not difficult to expand our concept of efficient resource allocations to include *technological progress*, the general term economists give to advances in the state of knowledge. In the present context, technological progress would be manifested as reductions over time in the cost of extraction. For a resource that can be extracted at constant marginal cost, a one-time breakthrough lowering the marginal cost of extraction would hasten the time of transition. Furthermore, for an increasing-cost resource, more of the total available resource would be recovered in the presence of technological progress than would be recovered without it. (Why?)

The most pervasive effects of technological progress involve continuous downward shifts in the cost of extraction over some time period. The total marginal cost of the resource could actually fall over time if the cost-reducing nature of technological progress became so potent that, in spite of increasing reliance on inferior ore, the marginal cost of extraction decreased (see Example 6.1). With a finite amount of this resource, the fall in total marginal cost would be transitory, since ultimately it would have to rise, but, as we shall see in the next few chapters, this period of transition can last quite a long time.

EXAMPLE 6.1

Historical Example of Technological Progress in the Iron Ore Industry

The term *technological progress* plays an important role in the economic analysis of mineral resources. Yet, at times, it can appear abstract, even mystical. It shouldn't! Far from being a blind faith detached from reality, technological progress refers to a host of ingenious ways in which people have reacted to impending shortages with sufficient imagination that the available supply of resources has been expanded by an order of magnitude and at reasonable cost. To illustrate how concrete a notion technological progress is, consider one example of how it has worked in the past.

In 1947 the president of Republic Steel, C. M. White, calculated the expected life of the Mesabi Range of northern Minnesota (the source of some 60 percent of iron ore consumed during World War II) as being in the range from five to seven years. By 1955, only eight years later, *U.S. News and World Report* concluded that worry over the scarcity of iron ore could be forgotten. The source of this remarkable transformation of a problem of scarcity into one of abundance was the discovery of a new technique of preparing iron ore, called *pelletization*.

Prior to pelletization, the standard ores from which iron was derived contained from 50 to more than 65 percent iron in crude form. There was a significant percentage of taconite ore available containing less than 30 percent iron in crude form, but no one knew how to produce it at reasonable cost. Pelletization is a process by which these ores are processed and concentrated at the mine site prior to shipment to the blast furnaces. The advent of pelletization allowed the profitable use of the taconite ores.

While expanding the supply of iron ore, pelletization reduced its cost in spite of the inferior grade being used. There were several sources of the cost reduction. First, substantially less energy was used; the shift in ore technology toward pelletization produced net energy savings of 17 percent in spite of the fact that the pelletization process itself required more energy. The reduction came from the discovery that the blast furnaces could be operated much more efficiently using pelletization inputs. The process also reduced labor requirements per ton by some 8.2 percent while increasing the output of the blast furnaces. A blast furnace owned by Armco Steel in Middletown, Ohio, which had a rated capacity of approximately 1,500 tons of molten iron per day, was able, by 1960, to achieve production levels of 2,700–2,800 tons per day when fired with 90 percent pellets. Pellets nearly doubled the blast furnace productivity!

Sources: Peter J. Kakela, "Iron Ore: Energy Labor and Capital Changes with Technology." SCIENCE, Vol. 202, December 15, 1978, pp. 1151–1157; and Peter J. Kakela, "Iron Ore: From Depletion to Abundance." SCIENCE, Vol. 212, April 10, 1981, pp. 132–136.

Market Allocations of Depletable Resources

In the preceding sections, we have examined in detail how the efficient allocation of substitutable, depletable, and renewable resources over time would be defined in a variety of circumstances. We must now address the question of whether actual markets can be expected to produce an efficient allocation. Can the private market, involving millions of consumers and producers each reacting to his or her own unique preferences, ever result in a dynamically efficient allocation? Is profit maximization compatible with dynamic efficiency?

Appropriate Property Rights Structures

The most common misconception of those who believe that even a perfect market could never achieve an efficient allocation of depletable resources is based on the idea that producers want to extract and sell the resources as fast as possible, since that is how they derive the value from the resource. This misconception makes people see markets as myopic and unconcerned about the future.

As long as the property rights governing natural resources have the characteristics of exclusivity, transferability, and enforceability (Chapter 2), the markets in which those resources are bought and sold will not necessarily lead to myopic choices. When bearing the marginal user cost, the producer acts in an efficient manner. A resource in the ground has two potential sources of value to its owner: (1) a use value when it is sold (the only source considered by those diagnosing inevitable myopia) and (2) an asset value when it remains in the ground. As long as the price of a resource continues to rise, the resource in the ground is becoming more valuable. The owner of this resource accrues this capital gain, however, only if the resource is conserved. A producer who sells all resources in the earlier periods loses the chance to take advantage of higher prices in the future.

A profit-maximizing producer attempts to balance present and future production in order to maximize the value of the resource. Since higher prices in the future provide an incentive to conserve, a producer who ignores this incentive would not be maximizing the value of the resource. We would expect resources owned by a myopic producer to be bought by someone willing to conserve and prepared to maximize its value. As long as social and private discount rates coincide, property rights structures are well defined, and reliable information about future prices is available, a producer who pursues maximum profits simultaneously provides the maximum present value of net benefits for society.

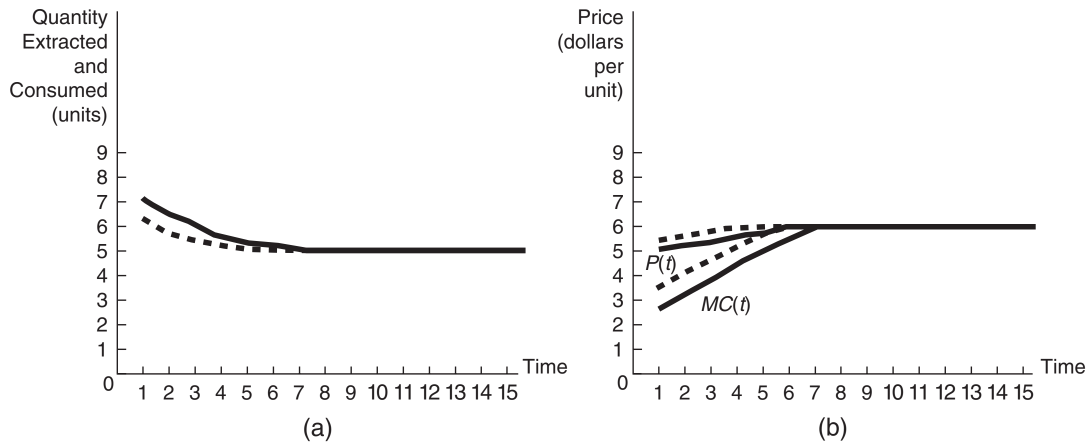
The implication of this analysis is that, in competitive resource markets, the price of the resource equals the total marginal cost of extracting and using the resource. Thus, Figures 6.2a through 6.5b can illustrate not only an efficient allocation but also the allocation produced by an efficient market. When used to describe an efficient market, the total marginal cost curve describes the time path that prices could be expected to follow.

Environmental Costs

One of the most important situations in which property rights structures may not be well defined is that in which the extraction of a natural resource imposes an environmental cost on society not internalized by the producers. The aesthetic costs of strip mining, the health risks associated with uranium tailings, and the acids leached into streams from mine operations are all examples of associated environmental costs. Not only is the presence of environmental costs empirically important, but also it is conceptually important, since it forms one of the bridges between the traditionally separate fields of environmental economics and natural resource economics.

Suppose, for example, that the extraction of the depletable resource caused some damage to the environment that was not adequately reflected in the costs faced by the extracting firms. This would be, in the context of discussion in Chapter 2, an external cost. The cost of getting the resource out of the ground, as well as processing and shipping it, is borne by the resource owner and considered in the calculation of how much of the resource to extract. The environmental damage, however, is not automatically borne by the owner and, in the absence of any outside attempt to internalize that cost, will not be part of the extraction decision. How would the market allocation, based on only the former cost, differ from the efficient allocation, which is based on both costs?

FIGURE 6.6 (a) Increasing Marginal Extraction Cost with Substitute Resource in the Presence of Environmental Costs: Quantity Profile (b) Increasing Marginal Extraction Cost with Substitute Resource in the Presence of Environmental Costs: Price Profile (Solid Line—without Environmental Costs; Dashed Line—with Environmental Costs)



We can examine this issue by modifying the numerical example used earlier in this chapter. Assume the environmental damage can be represented by increasing the marginal cost by \$1.[5] The additional dollar reflects the cost of the environmental damage caused by producing another unit of the resource. What effect do you think this would have on the efficient time profile for quantities extracted?

The answers are given in Figures 6.6a and 6.6b. The result of including environmental cost on the timing of the switch point is interesting because it involves two different effects that work in opposite directions. On the demand side, the inclusion of environmental costs results in higher prices, which tend to dampen demand. This lowers the rate of consumption of the resource, which, all other things being equal, would make it last longer.

All other things are not equal, however. The higher marginal cost also means that a smaller cumulative amount of the depletable resource would be extracted in an efficient allocation. (Why?) In our example shown in Figures 6.6a and 6.6b, the efficient cumulative amount extracted would be 30 units instead of the 40 units extracted in the case where environmental costs were not included. This supply-side effect tends to hasten the time when a switch to the renewable resource is made, all other things being equal.

[5]Including environmental damage, the marginal cost function would be raised to \$3 + $0.1Q$ instead of \$2 + $0.1Q$.

Which effect dominates—the rate of consumption effect or the supply effect? In our numerical example, the supply-side effect dominates and, as a result, the time of transition for an efficient allocation is sooner than for the market allocation. In general, the answer depends on the shape of the marginal-extraction-cost function. With constant marginal cost, for example, there would be no supply-side effect and the market would transition later. If the environmental costs were associated with the cost of the renewable resource, rather than the depletable resource, the time of transition for the efficient allocation would have been later than the market allocation. Can you see why?

What can we learn from these graphs about the allocation of depletable resources over time when environmental side effects are not borne by the agent determining the extraction rate? Ignoring external costs leaves the price of the depletable resource too low and too much of the resource would be extracted. This once again shows the interdependencies among the various decisions we have to make about the future. Environmental and natural resource decisions are intimately and inextricably linked.

Summary

The efficient allocation of depletable and renewable resources depends on the circumstances. When the resource can be extracted at a constant marginal cost, the efficient quantity of the depletable resource extracted declines over time. If no substitute is available, the quantity declines smoothly to zero. If a renewable constant-cost substitute is available, the quantity of the depletable resource extracted will decline smoothly to the quantity available from the renewable resource. In each case, all of the available depletable resource would be eventually used up and marginal user cost would rise over time, reaching a maximum when the last unit of depletable resource was extracted.

The efficient allocation of an increasing marginal-cost resource is similar in that the quantity extracted declines over time, but differs with respect to the behavior of marginal user cost and the cumulative amount extracted. Whereas marginal user cost typically rises over time when the marginal cost of extraction is constant, it declines over time when the marginal cost of extraction rises. Furthermore, in the constant-cost case the cumulative amount extracted is equal to the available supply; in the increasing-cost case it depends on the marginal extraction cost function.

Introducing technological progress and exploration activity into the model tends to delay the transition to renewable resources. Exploration expands the size of current reserves, while technological progress keeps marginal extraction cost from rising as much as it otherwise would. If these effects are sufficiently potent, marginal cost could actually decline for some period of time, causing the quantity extracted to rise.

When property rights structures are properly defined, market allocations of depletable resources can be efficient. Self-interest and efficiency are not inevitably incompatible.

When the extraction of resources imposes an external environmental cost, however, generally market allocations will not be efficient. The market price of the depletable resource would be too low, and too much of the resource would be extracted.

In an efficient market allocation, the transition from depletable to renewable resources is smooth and exhibits no overshoot-and-collapse characteristics. Whether the actual market allocations of these various types of resources are efficient remains to be seen. To the extent markets negotiate an efficient transition, a laissez-faire policy would represent an appropriate response by the government. On the other hand, if the market is not capable of yielding an efficient allocation, then some form of government intervention may be necessary. In the next few chapters, we shall examine these questions for a number of different types of depletable and renewable resources.

Discussion Question

1. One current practice is to calculate the years remaining for a depletable resource by taking the prevailing estimates of current reserves and dividing it by current annual consumption. How useful is that calculation? Why?

Self-Test Exercises

1. To anticipate subsequent chapters where more complicated renewable resource models are introduced, consider a slight modification of the two-period depletable resource model. Suppose a biological resource is renewable in the sense that any of it left unextracted after Period 1 will grow at rate k. Compared to the case where the total amount of a constant-*MEC* resource is fixed, how would the efficient allocation of this resource over the two periods differ? (*Hint*: It can be shown that $MNB_1/MNB_2 = (1 + k)/(1 + r)$, where *MNB* stands for marginal net benefit.)
2. Consider an increasing marginal-cost depletable resource with no effective substitute. (a) Describe, in general terms, how the marginal user cost for this resource in the earlier time periods would depend on whether the demand curve for that resource was stable or shifting outward over time. (b) How would the allocation of that resource over time be affected?
3. Many states are now imposing severance taxes on resources being extracted within their borders. In order to understand the effect of these on the allocation of the mineral over time, assume a stable demand curve. (a) How would the competitive allocation of an increasing marginal-cost depletable resource be affected by the imposition of a per-unit tax (e.g., $4 per ton) if there exists a constant-marginal-cost substitute? (b) Comparing the allocation without a

tax to one with a tax, in general terms, what are the differences in cumulative amounts extracted and the price paths?

4. For the increasing marginal-extraction-cost model of the allocation of a depletable resource, how would the ultimate cumulative amount taken out of the ground be affected by (a) an increase in the discount rate, (b) the extraction by a monopolistic, rather than a competitive, industry, and (c) a per-unit subsidy paid by the government for each unit of the abundant substitute used?
5. Suppose you wanted to hasten the transition from a depletable fossil fuel to solar energy. Compare the effects of a per-unit tax on the depletable resource to an equivalent per-unit subsidy on solar energy. Would they produce the same switch point? Why or why not?

Further Reading

Bohi, Douglas R., and Michael A. Toman. *Analyzing Nonrenewable Resource Supply* (Washington, DC: Resources for the Future, 1984). A reinterpretation and evaluation of research that attempts to weave together theoretical, empirical, and practical insights concerning the management of depletable resources.

Chapman, Duane. "Computation Techniques for Intertemporal Allocation of Natural Resources," *American Journal of Agricultural Economics* Vol. 69 (February 1987): 134–142. Shows how to find numerical solutions for the types of depletable resource problems considered in this chapter.

Conrad, Jon M., and Colin W. Clark. *Natural Resource Economics: Notes and Problems* (Cambridge: Cambridge University Press, 1987). Reviews techniques of dynamic optimization and shows how they can be applied to the management of various resource systems.

Toman, Michael A. "'Depletion Effects' and Nonrenewable Resource Supply," *Land Economics* Vol. 62 (November 1986): 341–353. An excellent, nontechnical discussion of the increasing-cost case with and without exploration and additions to reserves.

Additional References and Historically Significant References are available on this book's Companion Website: http://www.pearsonhighered.com/tietenberg/

Appendix

Extensions of the Constant Extraction Cost Depletable Resource Model: Longer Time Horizons and the Role of an Abundant Substitute

In the appendix to Chapter 5, we derived a simple model to describe the efficient allocation of a constant-marginal-cost depletable resource over time and presented the numerical solution for a two-period version of that model. In this appendix, the mathematical derivations for the extension to that basic model will be documented, and the resulting numerical solutions for these more complicated cases will be explained.

The *N*-Period, Constant-Cost, No-Substitute Case

The first extension involves calculating the efficient allocation of the depletable resource over time when the number of time periods for extraction is unlimited. This is a more difficult calculation because how long the resource will last is no longer predetermined; the time of exhaustion must be derived as well as the extraction path prior to exhaustion of the resource.

The equations describing the allocation that maximizes the present value of net benefits derived in the appendix to Chapter 3 are

$$\frac{a - bq_t - c}{(1 + r)^{t-1}} - \lambda = 0, t = 1, \ldots, T \tag{1}$$

$$\sum_{t=1}^{T} q_t = \bar{Q} \tag{2}$$

The parameter values assumed for the numerical example presented in the text are

$$a = \$8, b = 0.4, c = \$2, \bar{Q} = 40, \text{ and } r = 0.10$$

The allocation that satisfies these conditions is

$q_1 = 8.004$	$q_4 = 5.689$	$q_7 = 2.607$	$T = 9$
$q_2 = 7.305$	$q_5 = 4.758$	$q_8 = 1.368$	$\lambda = 2.7983$
$q_3 = 6.535$	$q_6 = 3.733$	$q_9 = 0.000$	

The optimality of this allocation can be verified by substituting these values into the above equations. (Due to rounding, these add to 39.999, rather than 40.000.)

Practically speaking, solving these equations to find the optimal solution is not a trivial matter, but neither is it very difficult. One method of finding the solution for those without the requsite mathematics involves developing a computer algorithm (computation procedure) that converges on the correct answer. One such algorithm for this example can be constructed as follows: (1) assume a value for λ; (2) using Equation set (1) solve for all q's based upon this λ; (3) if the sum of the calculated q's exceeds $\bar{Q}$, adjust λ upward or if the sum of the calculated q's is less than $\bar{Q}$, adjust λ downward (the adjustment should use information gained in previous steps to ensure that the new trial will be closer to the solution value); (4) repeat steps (2) and (3) using the new λ; (5) when the sum of the q's is sufficiently close to $\bar{Q}$ stop the calculations. As an exercise, those interested in computer programming might construct a program to reproduce these results.

Constant Marginal Cost with an Abundant Renewable Substitute

The next extension assumes the existence of an abundant, renewable, perfect substitute, available in unlimited quantities at a cost of \$6 per unit. To derive the dynamically efficient allocation of both the depletable resource and its substitute, let q_t be the amount of a constant-marginal-cost depletable resource extracted in year t and q_{st} the amount used of another constant-marginal-cost resource that is perfectly substitutable for the depletable resource. The marginal cost of the substitute is assumed to be \$$d$.

With this change, the total benefit and cost formulas become

$$\text{Total benefit} = \sum_{t=1}^{T} a(q_t + q_{st}) - \frac{b}{2}(q_t + q_{st})^2 \tag{3}$$

$$\text{Total cost} = \sum_{t=1}^{T} (cq_t + dq_t) \tag{4}$$

The objective function is thus

$$PVNB = \sum_{t=1}^{T} \frac{a(q_t + q_{st}) - \frac{b}{2}(q_t^2 + q_{st}^2 + 2q_t q_{st}) - cq_t - dq_{st}}{(1 + r)^{t-1}} \tag{5}$$

subject to the constraint on the total availability of the depletable resource

$$\bar{Q} - \sum_{t=1}^{T} q_t \geq 0 \tag{6}$$

Necessary and sufficient conditions for an allocation maximizing this function are expressed in Equations (7), (8), and (9):

$$\frac{a - b(q_t + q_{st}) - c}{(1 + r)^{t-1}} - \lambda \leq 0, t = 1, \ldots, T \tag{7}$$

Any member of Equation set (7) will hold as an equality when $q_t > 0$ and will be negative when

$$a - b(q_t + q_{st}) - d \leq 0, t = 1, \ldots, T \tag{8}$$

Any member of Equation set (8) will hold as an equality when $q_{st} > 0$ and will be negative when $q_{st} = 0$

$$\bar{Q} - \sum_{t=1}^{T} q_t \geq 0 \tag{9}$$

For the numerical example used in the test, the following parameter values were assumed: a = \$8, b = 0.4, c = \$2, d = \$6, Q = 40, and r = 0.10. It can be readily verified that the optimal conditions are satisfied by

$$\begin{aligned}
&q_1 = 8.798 \quad q_3 = 7.495 \quad q_5 = 5.919 \\
&q_2 = 8.177 \quad q_4 = 6.744 \\
&q_{s6} = 2.137 \quad q_{st} = \begin{cases} 5.000 \text{ for } t>6 \\ 0 \text{ for } t<6 \end{cases} \\
&q_6 = 2.863 \quad \lambda = 2.481
\end{aligned}$$

The depletable resource is used up before the end of the sixth period and the switch is made to the substitute resource at that time. From Equation set (8), in competitive markets the switch occurs precisely at the moment when the resource price rises to meet the marginal cost of the substitute.

The switch point in this example is earlier than in the previous example (the sixth period rather than the ninth period). Since all characteristics of the problem except for the availability of the substitute are the same in the two numerical examples, the difference can be attributed to the availability of the renewable substitute.

7 Energy: The Transition from Depletable to Renewable Resources

If it ain't broke, don't fix it!

—Old Maine proverb

Introduction

Energy is one of our most critical resources; without it, life would cease. We derive energy from the food we eat. Through photosynthesis, the plants we consume—both directly and indirectly when we eat meat—depend on energy from the sun. The materials we use to build our houses and produce the goods we consume are extracted from the earth's crust, and then transformed into finished products with expenditures of energy.

Currently, many industrialized countries depend on oil and natural gas for the majority of their energy needs. According to the International Energy Agency (IEA), these resources together supply 59 percent of all primary energy consumed worldwide. (Adding coal, another fossil fuel resource, increases the share to 86 percent of the total.) Fossil fuels are depletable, nonrecyclable sources of energy. Crude oil proven reserves peaked during the 1970s and natural gas peaked in the 1980s in the United States and Europe, and since that time, the amount extracted has exceeded additions to reserves.

Kenneth Deffeyes (2001) and Campbell and Laherrere (1998) estimate that *global* oil production will peak early in the twenty-first century. As Example 7.1 points out, however, due to the methodology used, these predictions of the timing of the peak are controversial.

Even if we cannot precisely determine when the fuels on which we currently depend so heavily will run out, we still need to think about the process of transition to new energy sources. According to depletable resource models, oil and natural gas would be used until the marginal cost of further use exceeded the marginal cost of substitute resources—either more abundant depletable resources such as coal, or renewable sources such as solar energy.[1] In an efficient market path, the transition to these alternative sources would be smooth and harmonious. Have the allocations of the last several decades been efficient or not? Is the market mechanism flawed in

[1]When used for other purposes, oil can be recyclable. Waste lubricating oil is now routinely recycled.

EXAMPLE 7.1

Hubbert's Peak

When can we expect to run out of oil? It's a simple question with a complex answer. In 1956 geophysicist M. King Hubbert, then working at the Shell research lab in Houston, predicted that U.S. oil production would reach its peak in the early 1970s. Though Hubbert's analysis failed to win much acceptance from experts either in the oil industry or among academics, his prediction came true in the early 1970s. With some modifications, this methodology has since been used to predict the timing of a downturn in global annual oil production as well as when we might run out of oil.

These forecasts and the methods that underlie them are controversial, in part because they ignore such obvious economic factors as prices. The Hubbert model assumes that the annual rate of production follows a bell-shaped curve, regardless of what is happening in oil markets; oil prices don't matter. It seems reasonable to believe, however, that by affecting the incentive to explore new sources and to bring them into production, prices should affect the shape of the production curve.

How much difference would incorporating prices make? Pesaran and Samiei (1995) find, as expected, that modifying the model to include price effects causes the estimated ultimate resource recovery to be larger than implied by the basic Hubbert model. Moreover, a study by Kaufman and Cleveland (2001) finds that forecasting with a Hubbert-type model is fraught with peril.

> *. . . production in the lower 48 states stabilizes in the late 1970's and early 1980's, which contradicts the steady decline forecast by the Hubbert model. Our results indicate that Hubbert was able to predict the peak in US production accurately because real oil prices, average real cost of production, and [government decisions] co-evolved in a way that traced what appears to be a symmetric bell-shaped curve for production over time. A different evolutionary path for any of these variables could have produced a pattern of production that is significantly different from a bell-shaped curve and production may not have peaked in 1970. In effect, Hubbert got lucky. [p. 46]*

Does this mean we are not running out of oil? No. It simply means we have to be cautious when interpreting forecasts of the timing of the transition to other sources of energy. In 2005, the Administrator of the U.S. Energy Information Agency (EIA) presented a compendium of 36 studies of global oil production and all but one forecasted a production peak. The EIA's own estimates of the timing range from 2031 to 2068 (Caruso, 2005). The issue, it seems, is no longer whether oil production will peak, but when.

Sources: M. Pesaran and H. Samiei, "Forecasting Ultimate Resource Recovery." INTERNATIONAL JOURNAL OF FORECASTING, Vol. 11, No. 4 (1995), pp. 543–555; R. Kaufman and C. Cleveland, "Oil Production in the Lower 48 States: Economic, Geological, and Institutional Determinants." ENERGY JOURNAL, Vol. 22, No. 1 (2001), pp. 27–49; and G. Caruso, "When Will World Oil Production Peak?" A presentation at the 10th Annual Asia Oil and Gas Conference in Kuala Lumpur, Malaysia, June 13, 2005.

its allocation of depletable, nonrecyclable resources? If so, is it a fatal flaw? If not, what caused the inefficient allocations? Is the problem correctable?

In this chapter we shall examine some of the major issues associated with the allocation of energy resources over time and explore how economic analysis can clarify our understanding of both the sources of the problems and their solutions.

Natural Gas: Price Controls

In the United States, during the winter of 1974 and early 1975, serious shortages of natural gas developed. Customers who had contracted for and were willing to pay for natural gas were unable to get as much as they wanted. The shortage (or curtailments, as the Federal Energy Regulatory Commission (FERC) calls them) amounted to two trillion cubic feet of natural gas in 1974–1975, which represented roughly 10 percent of the marketed production in 1975. In an efficient allocation, shortages of that magnitude would never have materialized. What happened?

The source of the problem can be traced directly to government controls over natural gas prices. This story begins, oddly enough, with the rise of the automobile, which traditionally has not used natural gas as a fuel. The increasing importance of the automobile for transportation created a rising demand for gasoline, which in turn stimulated a search for new sources of crude oil. This exploration activity uncovered large quantities of natural gas (known as associated gas), in addition to large quantities of crude oil, which was the object of the search.

As natural gas was discovered, it replaced manufactured gas—and some coal—in the geographic areas where it was found. Then, as a geographically dispersed demand developed for this increasingly available gas, a long-distance system of gas pipelines was constructed. In the period following World War II, natural gas became an important source of energy for the United States.

The regulation of natural gas began in 1938 with the passage of the Natural Gas Act. This act transformed the Federal Power Commission (FPC), which subsequently become FERC, into a federal regulatory agency charged with maintaining "just" prices. In 1954 a Supreme Court decision in *Phillips Petroleum Co. v. Wisconsin* forced the FPC to extend their price control regulations to the producer. Previously they had limited their regulation to pipeline companies.

Because the process of setting price ceilings proved cumbersome, the hastily conceived initial "interim" ceilings remained in effect for almost a decade before the Commission was able to impose more carefully considered ceilings. What was the effect of this regulation?

By returning to our models in the previous section, we can see the havoc this would raise. The ceiling would prevent prices from reaching their normal levels. Since price increases are the source of the incentive to conserve, the lower prices would cause more of the resource to be used in earlier years. Consumption levels in those years would be higher under price controls than without them.

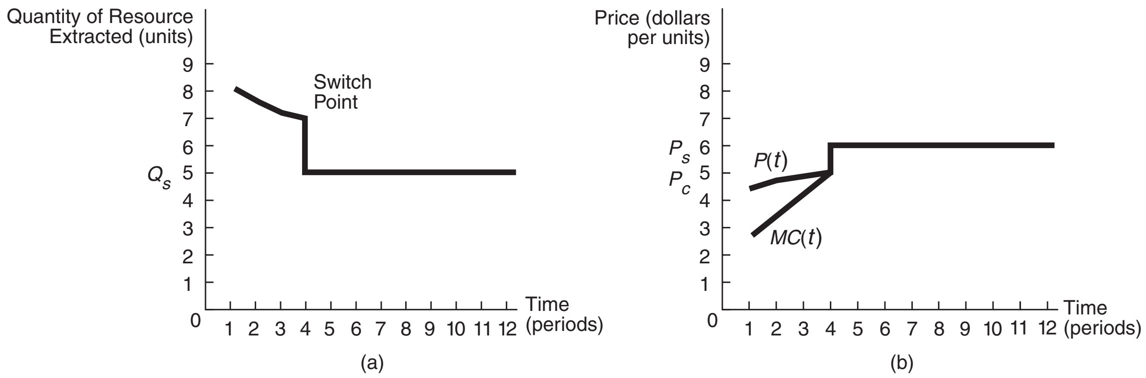
Effects on the supply side are also significant. Producers would produce the resource only when they could do so profitably. Once the marginal cost rose to meet the price ceiling, no more would be produced, in spite of the large demand for the resource at that price. Thus, as long as price controls were permanent, less of the resource would be produced with controls than without. Furthermore, more of what would be produced would be used in the earlier years.

The combined impact of these demand-and-supply effects would be to distort the allocation significantly (see Figures 7.1a and 7.1b). While a number of aspects differentiate this allocation from an efficient one, several are of particular importance: (1) more of the resource is left in the ground, (2) the rate of consumption is too high, (3) the time of transition is earlier, and (4) the transition is abrupt, with prices suddenly jumping to new, higher levels. All are detrimental. The first effect means we would not be using all of the natural gas available at prices consumers were willing to pay. Because price controls would cause prices to be lower than efficient, the resource would be depleted too fast. These two effects would cause an earlier and abrupt transition to the substitute possibly before the technologies to use it were adequately developed.

The discontinuous jump to a new technology, which results from price controls, can place quite a burden on consumers. Attracted by artificially low prices, consumers would invest in equipment to use natural gas, only to discover—after the transition—that natural gas was no longer available.

One interesting characteristic of price ceilings is that they affect behavior even before they are binding.[2] This effect is clearly illustrated in Figures 7.1a and 7.1b in the earlier years. Even though the price in the first year is lower than the price ceiling, it is not equal to the efficient price. (Can you see why? Think what price controls do to the marginal user cost faced by producers.) The price ceiling causes a reallocation of resources toward the present, which, in turn, affects prices in the earlier years.

FIGURE 7.1 (a) Increasing Marginal Extraction Cost with Substitute Resource in the Presence of Price Controls: Quantity Profile (b) Increasing Marginal Extraction Cost with Substitute Resource in the Presence of Price Controls: Price Profile



[2]For a complete analysis of this point, see Lee (1978).

Price controls may cause other problems as well. Up to this point, we have discussed permanent controls. Not all price controls are permanent; they can change unpredictably at the whim of the political process. The fact that prices could suddenly rise when the ceiling is lifted also creates unfortunate incentives. If producers expect a large price increase in the near future, they have an incentive to stop production and wait for the higher prices. This could cause severe problems for consumers.

For legal reasons, the price controls on natural gas were placed solely on gas shipped across state lines. Gas consumed within the states where it was produced could be priced at what the market would bear. As a result, gas produced and sold within the state received a higher price than that sold in other states. Consequently, the share of gas in the interstate market fell over time as producers found it more profitable to commit reserve additions to the *intrastate*, rather than the *interstate*, market. During 1964–1969, about 33 percent of the average annual reserve additions were committed to the interstate market. By 1970–1974, this commitment had fallen to a little less than 5 percent.

The practical effect of forcing lower prices for gas destined for the interstate market was to cause the shortages to be concentrated in states served by pipeline and dependent on the interstate shipment of gas. As a result, the ensuing damage was greater than it would have been if all consuming areas had shared somewhat more equitably in the shortfall. Governmental control of prices not only precipitated the damage, it intensified it!

It seems fair to conclude that, by sapping the economic system of its ability to respond to changing conditions, price controls on natural gas created a significant amount of turmoil. If this kind of political control is likely to recur with some regularity, perhaps the overshoot and collapse scenario might have some validity. In this case, however, it would be caused by government interference rather than any pure market behavior. If so, the proverb that opens this chapter becomes particularly relevant!

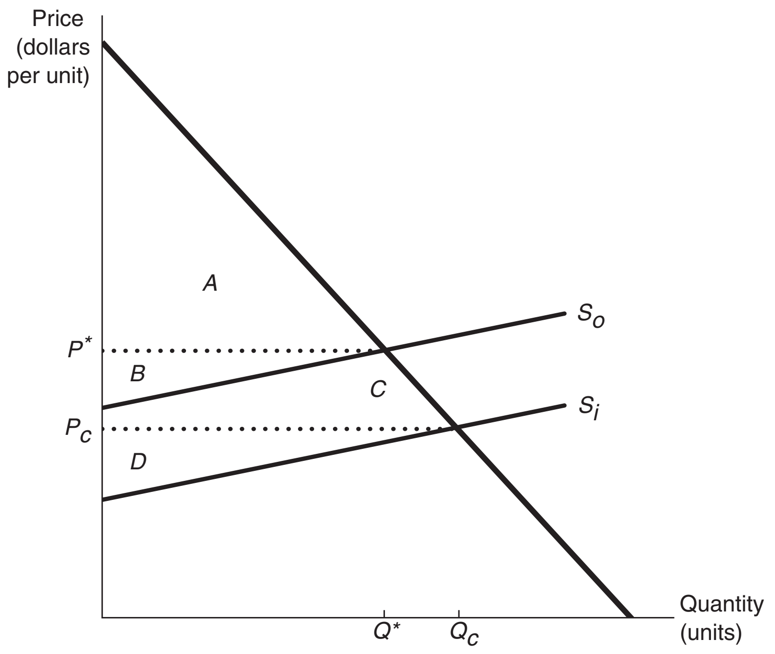
Why did Congress embark on such a counterproductive policy? The answer is found in rent-seeking behavior that can be explained through the use of our consumer and producer surplus model. Let's examine the political incentives in a simple model.

Consider Figure 7.2. An efficient market allocation would result in Q^* supplied at price P^*. The net benefits received by the country would be represented by areas A and B. Of these net benefits, area A would be received by consumers as consumer surplus and B would be received by producers as a producer surplus.

Suppose now that a price ceiling were established. From the above discussion we know that this ceiling would reduce the marginal user cost because higher future prices would no longer be possible. In Figure 7.2, this has the effect for current producers of lowering the perceived supply curve, due to the lower marginal user cost. As a result of this shift in the perceived supply curve, current production would expand to Q_c and price would fall to P_c. Current consumers would unambiguously be better off, since consumer surplus would be area $A + B + C$ instead of area A. They would have gained a net benefit equal to $B + C$.

It may appear that producers could also gain if $D > B$, but that is not correct because this diagram does not take into account the effects on other time periods.

FIGURE 7.2 The Effect of Price Controls



Since producers would be overproducing, they would be giving up the scarcity rent they could have gotten without price controls. Area *D* measures current profits only without considering scarcity rent. When the loss in scarcity rent is considered, producers unambiguously lose net benefits.

Future consumers are also unambiguously worse off. In the terms of Figure 7.2, which represents the allocation in a given year, as the resource was depleted, the supply curve for each subsequent year would shift up, thereby reflecting the higher marginal extraction costs for the remaining endowment of the resource. When the marginal extraction cost ultimately reached the level of the price control, the amount supplied would drop to zero. Extracting more would make no sense to suppliers because their cost would exceed the controlled price. Since the demand would not be zero at that price, a shortage would develop. Although consumers would be willing to pay higher prices and suppliers would be happy to supply more of the resource at those higher prices (if they were not prevented from doing so by the price control), the price ceiling would keep those resources in the ground.

Congress may view scarcity rent as a possible source of revenue to transfer from producers to consumers. As we have seen, however, scarcity rent is an opportunity cost that serves a distinct purpose—the protection of future consumers. When a government attempts to reduce this scarcity rent through price controls, the result is an overallocation to current consumers and an underallocation to future consumers. Thus, what appears to be a transfer from producers to consumers is, in large part, also a transfer from future consumers to present consumers. Since current consumers mean current votes and future consumers may not know whom to blame by the time shortages appear, price controls are politically attractive. Unfortunately, they are also inefficient; the losses to future consumers and producers are greater than the gains to current consumers. Because controls distort the allocation toward the present, they are also unfair. Thus, markets in the presence of price controls are indeed myopic, but the problem lies with the controls, not the market.

Over the long run, price controls end up harming consumers rather than helping them. Scarcity rent plays an important role in the allocation process, and attempts to eliminate it can create more problems than solutions. After long debating the price control issue, Congress passed the Natural Gas Policy Act on November 9, 1978. This act initiated the eventual phased decontrol of natural gas prices. By January 1993, no sources of natural gas were subject to price controls.

Oil: The Cartel Problem

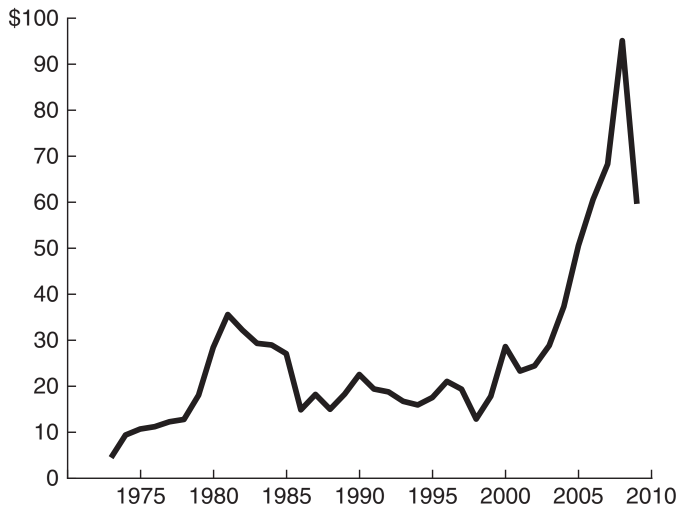
Since we have considered similar effects on natural gas, we note merely that historically price controls have been responsible for much mischief in the oil market as well. A second source of misallocation in the oil market, however, deserves further consideration. Most of the world's oil is produced by a cartel called the Organization of Petroleum Exporting Countries (OPEC). The members of this organization collude to exercise power over oil production and prices. As established in Chapter 2, seller power over resources due to a lack of effective competition leads to an inefficient allocation. When sellers have market power, they can restrict supply and thus force prices higher than otherwise. (Figure 7.3 shows oil prices over time.)

Though these conclusions were previously derived for nondepletable resources, they are valid for depletable resources as well. A monopolist can extract more scarcity rent from a depletable resource base than competitive suppliers simply by restricting supply. The monopolistic transition results in a slower rate of production and higher prices.[3] The monopolistic transition to a substitute, therefore, occurs later than a competitive transition. It also reduces the net present value society receives from these resources.

The cartelization of the oil suppliers has apparently been very effective (Smith, 2005). Why? Are the conditions that make it profitable unique to oil, or could oil cartelization be the harbinger of a wave of natural resource cartels? To answer these questions, we must isolate those factors that make oil cartelization possible. Although many factors are involved, four stand out: (1) the price elasticity of demand for OPEC oil in both the long run and the short run; (2) the income elasticity of demand for oil; (3) the supply responsiveness of the oil producers who are not OPEC members; and (4) the compatibility of interests among members of OPEC.

[3]The conclusion that a monopoly would extract a resource more slowly than a competitive mining industry is not perfectly general. It is possible to construct demand curves such that the extraction of the monopolist is greater than or equal to that of a competitive industry. As a practical matter, these conditions seem unlikely. That a monopoly would restrict output, while not inevitable, is the most likely outcome.

FIGURE 7.3 Real Crude Oil Price (1973–2009)



Sources: Monthly Energy Review (MER), U.S. Energy Information Administration (EIA) (http://www.eia.doe.gov/mer/prices.html); Consumer Price Index (CPI), Bureau of Labor Statistics (http://www.bls.gov/cpi/data.htm).

Note: Prices are in 2009 dollars.

Price Elasticity of Oil Demand

The elasticity of demand is an important ingredient because it determines how responsive demand is to price. When demand elasticities are between 0 and –1 (i.e., when the percent quantity response is smaller than the percent price response), price increases lead to increased revenue. Exactly how much revenue would increase when prices increase depends on the magnitude of the elasticity. Generally, the smaller the absolute value of the price elasticity of demand (the closer it is to 0.0), the larger the gains to be derived from forming a cartel.

The price elasticity of demand for oil depends on the opportunities for conservation, as well as on the availability of substitutes. As storm windows cut heat losses, the same temperature can be maintained with less heating oil. Smaller, more fuel-efficient automobiles reduce the amount of gasoline needed to travel a given distance. The larger the set of these opportunities and the smaller the cash outlays required to exploit them, the more price-elastic the demand. This suggests that demand will be more price-elastic in the long run (when sufficient time has passed to allow adjustments) than in the short run.

The availability of substitutes is also important because it limits the degree to which prices can be raised by a producer cartel. Abundant quantities of substitutes available at prices not far above competitive oil prices can set an upper limit on the cartel price. Unless OPEC controls those sources as well—and it doesn't—any attempts to raise prices above those limits would cause the consuming nations to simply switch to these alternative sources; OPEC would have priced itself out of the market.

As described in more detail in a subsequent section of this chapter, alternative sources clearly exist, although they are expensive and the time of transition is long.

Income Elasticity of Oil Demand

The income elasticity of oil demand is important because it indicates how sensitive oil demand is to growth in the world economy. As income grows, oil demand should grow. This continual increase in demand fortifies the ability of OPEC to raise its prices.

The income elasticity of demand is also important because it registers how sensitive demand is to the business cycle. The higher the income elasticity of demand, the more sensitive demand is to periods of rapid economic growth or to recessions. This sensitivity was a major source of the 1983 weakening of the cartel and the significant fall in oil prices starting in late 2008. A recession caused a large reduction in the demand for oil, putting new pressure on the cartel to absorb this demand reduction. Conversely, whenever the global economy recovers, the cartel benefits disproportionately.

Non-OPEC Suppliers

Another key factor in the ability of producer nations to exercise power over a natural resource market is their ability to prevent new suppliers, not part of the cartel, from entering the market and undercutting the price. Currently OPEC produces about 45 percent of the world's oil. If the remaining producers were able, in the face of higher prices, to expand their supply dramatically, they would increase the amount of oil supplied and cause the prices to fall, which would decrease OPEC's market share. If this response were large enough, the allocation of oil would approach the competitive allocation.

Recognizing this impact, the cartel must take the nonmembers into account when setting prices. Salant (1976) proposed an interesting model of monopoly pricing in the presence of a fringe of small nonmember producers that serves as a basis for exploring this issue. His model includes a number of suppliers. Some form a cartel. Others, a smaller number, form a "competitive fringe." The cartel is assumed to set the price of oil to maximize its collective profits by restricting its production, taking the competitive fringe production into account. The competitive fringe cannot directly set the price, but, since it is free to choose the level of production that maximizes its own profits, its output does affect the cartel's pricing strategy.

What conclusions does this model yield? The model concludes that a resource cartel would set different prices in the presence of a competitive fringe than in its absence. With a competitive fringe, it would set the initial price somewhat lower than the pure monopoly price and allow price to rise more rapidly. This strategy maximizes cartel profits by forcing the competitive fringe to produce more in the earlier periods (in response to higher demand) and eventually to exhaust their supplies. Once the competitive fringe has depleted its reserves, the cartel would raise the price and thereafter prices would increase much more slowly.

Thus, the optimal strategy, from the point of view of the cartel, is to hold back on its own sales during the initial period, letting the other suppliers exhaust their supplies. Sales and profits of the competitive fringe, in this optimal cartel strategy, decline over time, while sales and profits of the cartel increase over time as prices rise and the cartel captures a larger share of the market.

One fascinating implication of this model is that the formation of the cartel raises the present value of competitive fringe profits by an even greater percentage than the present value of cartel profits. Those without the power gain more in percentage terms than those with the power!

Though this may seem counterintuitive, it is actually easily explained. The cartel, in order to keep the price up, must cut back on its own production level. The competitive fringe, however, is under no such constraint and is free to take advantage of the high prices caused by the cartel's withheld production without cutting back its own production. Thus, the profits of the competitive fringe are higher in the earlier period, which, in present value terms, are discounted less. All the cartel can do is wait until the competitive suppliers become less of a force in the market. The implication of this model is that the competitive fringe is a collective force in the oil market, even if it controls as little as one-third of the production.

The impact of this competitive fringe on OPEC behavior was dramatically illustrated by events in the 1985–1986 period. In 1979, OPEC accounted for approximately 50 percent of world oil production, while in 1986 this had fallen to approximately 30 percent. Taking account of the fact that total world oil production was down during this period over 10 percent for all producers, the pressures on the cartel mounted and prices ultimately fell. The real cost of crude oil imports in the United States fell from $34.95 per barrel in 1981 to $11.41 in 1986. OPEC simply was not able to hold the line on prices because the necessary reductions in production were too large for the cartel members to sustain.

In the summer of 2008, the price of crude oil soared above $138 per barrel. The price increase was due to strong worldwide demand coupled with restricted supply from Iraq because of the war. However, these high prices also underscored the major oil companies' difficulty finding new sources outside of OPEC countries. High oil prices in the 1970s drove Western multinational oil companies away from low-cost Middle Eastern oil to high-cost new oil in places such as the North Sea and Alaska. Most of these oil companies are now running out of big non-OPEC opportunities, which diminishes their ability to moderate price.

Compatibility of Member Interests

The final factor we shall consider in determining the potential for cartelization of natural resource markets is the internal cohesion of the cartel. With only one seller, the objective of that seller can be pursued without worrying about alienating others who could undermine the profitability of the enterprise. In a cartel composed of many sellers, that freedom is no longer as wide ranging. The incentives of each member and the incentives of the group as a whole may diverge.

Cartel members have a strong incentive to cheat. A cheater, if undetected by the other members, could surreptitiously lower its price and steal part of the market from the others. Formally, the price elasticity of demand facing an individual member is substantially higher than that for the group as a whole, because some of the increase in individual sales at a lower price represents sales reductions for other members. When producers face markets characterized by a high price elasticity, lower prices maximize profits. Thus, successful cartelization presupposes a means for detecting cheating and enforcing the collusive agreement.

In addition to cheating, however, cartel stability is also threatened by the degree to which members fail to agree on pricing and output decisions. Oil provides an excellent example of how these dissensions can arise. Since the 1974 rise of OPEC as a world power, Saudi Arabia has frequently exercised a moderating influence on the pricing decisions of OPEC. Why?

One highly significant reason is the size of Saudi Arabia's oil reserves (see Table 7.1). Saudi Arabia's reserves are larger than those of any other member. Hence, Saudi Arabia has an incentive to preserve the value of those resources. Setting prices too high would undercut the future demand for its oil. As previously stated, the demand for oil in the long run is more price-elastic than in the short run. Countries with smaller reserves, such as Nigeria, know that in the long run their reserves will be gone and therefore these countries are more concerned about the near future. Since alternative sources of supply are not much of a threat in the near future because of long development times, countries with small reserves want to extract as much rent as possible now.

TABLE 7.1 The World's Largest Oil Reserves

Country	Reserves (in billions of barrels)
Saudi Arabia	266.7
Canada[1]	178.1
Iran	136.2
Iraq	115.0
Kuwait	104.0
Venezuela	99.0
United Arab Emirates	97.8
Russia	60.0
Libya	43.7
Nigeria	36.2
Kazakhstan	30.0
United States	21.3

[1] PennWell Corporation, *Oil & Gas Journal*, Vol. 106, No. 48 (December 22, 2008), except United States. Oil includes crude oil and condensate. Data for the United States are from the Energy Information Administration, *U.S. Crude Oil, Natural Gas, and Natural Gas Liquids Reserves, 2007 Annual Report*, DOE/EIA-0216(2007) (February 2009). *Oil & Gas Journal*'s oil reserve estimate for Canada includes 5.392 billion barrels of conventional crude oil and condensate reserves and 172.7 billion barrels of oil sands reserves.

Source: http://www.eia.doe.gov/emeu/international/oilreserves.html compiled from PennWell Corporation, *Oil & Gas Journal*, Vol. 106, No. 48 (December 22, 2008).

The size of Saudi Arabia's production not only provides an incentive to preserve the stability of the oil market over the longer run, it also gives it the potential to make its influence felt. Its capacity to produce is so large that it can unilaterally affect world prices as long as it has excess capacity to use in pursuit of this goal.

This examination of the preconditions for successful cartelization reveals that creating a successful cartel is not an easy path to pursue for producers. It is therefore not surprising that OPEC has had its share of trouble exercising control over price in the oil market. When possible, however, cartelization can be very profitable. When the resource is a strategic raw material on which consuming nations have become dependent, cartelization can be very costly to those consuming nations.

Strategic-material cartelization also confers on its members political, as well as economic, power. Economic power can become political power when the revenue is used to purchase weapons or the capacity to produce weapons. The producer nations can also use an embargo of the material as a lever to cajole reluctant adversaries into foreign policy concessions. When the material is of strategic importance, the potential for embargoes casts a pall over the normally clear and convincing case for free trade of raw materials among nations.

Fossil Fuels: National Security and Climate Considerations

The Climate Dimension

All fossil fuels contain carbon. When these fuels are burned, unless the resulting carbon is captured, it is released into the atmosphere as carbon dioxide. As explained in more detail in Chapter 16, carbon dioxide is a greenhouse gas, which means that it is a contributor to what is known popularly as global warming, or more accurately (since the changes are more complex than simply universal warming) as climate change.

Climate considerations affect energy policy in two ways: (1) the level of energy consumption matters (as long as carbon-emitting sources are part of the mix) and (2) the mix of energy sources matters (since some emit more carbon than others). As can be seen from Table 7.2, among the fossil fuels, coal contains the most carbon per unit of energy produced and natural gas contains the least.

From an economic point of view, the problem with how the market makes energy choices is that in the absence of explicit regulation, emissions of carbon generally involve an externality to the energy user. Therefore, we would expect that market choices, which are based upon the relative private costs of using these fuels, would involve an inefficient bias toward fuels containing carbon, thereby jeopardizing the timing and smooth transition toward fuels that pose less of a climate change threat. In Chapter 16, we shall cover a host of policies that can be used to internalize those costs, but in the absence of those policies it might be necessary to subsidize renewable sources of energy that have little or no carbon.

TABLE 7.2 Carbon Content of Fuels

Fuel Type	Metric Tons of Carbon (per billion BTUs)
Coal	25.61
Coal (Electricity Generation)	25.71
Natural Gas	14.47
Residual Fuel Oil	21.49
Oil (Electricity Generation)	19.95
Liquefied Petroleum Gas	17.02
Distillate Fuel Oil	19.95

Source: Energy Information Administration.

The National Security Dimension

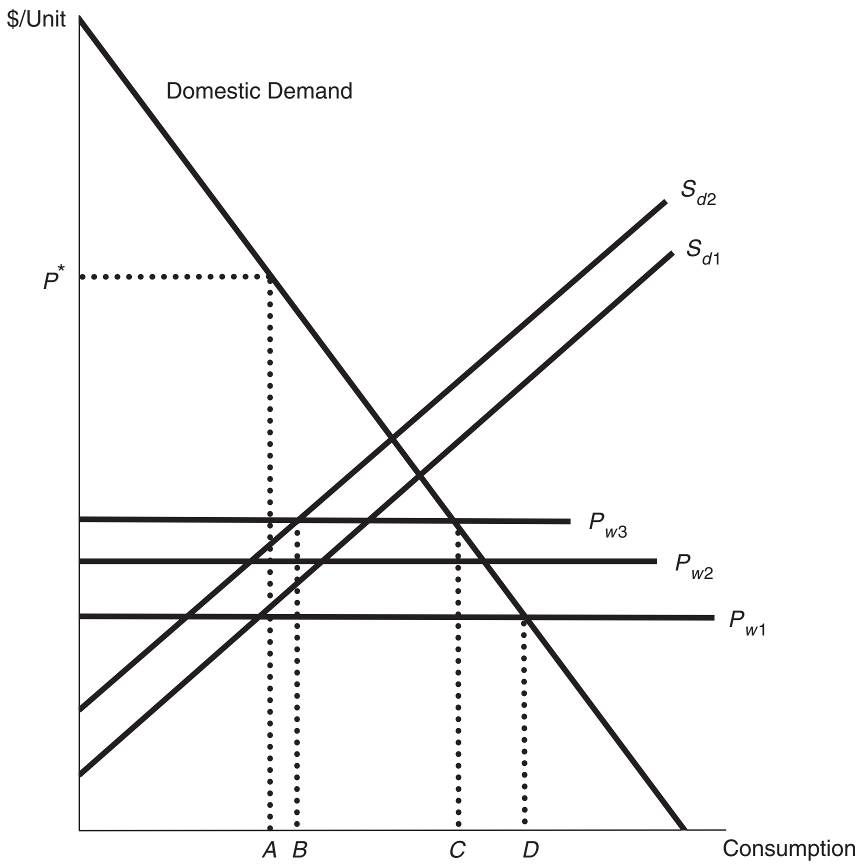
Vulnerable strategic imports also have an added cost that is not reflected in the marketplace. National security is a classic public good. No individual importer correctly represents our collective national security interests in making a decision on how much to import. Hence, leaving the determination of the appropriate balance between imports and domestic production to the market generally results in an excessive dependence on imports due to both climate change and national security considerations (see Figure 7.4).

In order to understand the interaction of these factors five supply curves are relevant. Domestic supply is reflected by two options. The first, S_{d1}, is the long-run domestic supply curve without considering the climate change damages resulting from burning more oil, while the second, S_{d2}, is the domestic supply curve that includes these per-unit damages. Their upward slopes reflect increasing availability of domestic oil at higher prices, given sufficient time to develop those resources. Imported foreign oil is reflected by three supply curves: P_{w1} reflects the observed world price, P_{w2} includes a "vulnerability premium" in addition to the world price, and P_{w3} adds in the per-unit climate change damages due to consuming more imported oil. The vulnerability premium reflects the additional national security costs caused by imports. All three curves are drawn horizontally to the axis to reflect the assumption that any importing country's action on imports is unlikely to affect the world price for oil.

As shown in Figure 7.4, in the absence of any correction for national security and climate change considerations, the market would generally demand and receive *D* units of oil. Of this total amount, *A* would be domestically produced and *D-A* would be imported. (Why?)

In an efficient allocation, incorporating the national security and climate change considerations, only *C* units would be consumed. Of these, *B* would be domestically produced and *C-B* would be imported. Note that because national security and climate change are externalities, the market in general tends to consume too much oil and vulnerable imports exceed their efficient level.

FIGURE 7.4 The National Security Problem



What would happen during an embargo? Be careful! At first glance, you would guess that we would consume where domestic supply equals domestic demand, but that is not right. Remember that S_{d1} is the domestic supply curve, *given enough time to develop the resources*. If an embargo hits, developing additional resources cannot happen immediately (multiple year time lags are common). Therefore, in the short run, the supply curve becomes perfectly inelastic (vertical) at *A*. The price will rise to P^* to equate supply and demand. As the graph indicates, the loss in consumer surplus during an embargo can be very large indeed.

How can importing nations react to this inefficiency? As Debate 7.1 shows, several strategies are available.

The importing country might be able to become self-sufficient, but should it? If the situation is adequately represented by Figure 7.4, then the answer is clearly no. The net benefit from self-sufficiency (the allocation where domestic supply S_{d1} crosses the demand curve) is clearly lower than the net benefit from the efficient allocation (*C*).

DEBATE 7.1

How Should the United States Deal with the Vulnerability of Its Imported Oil?

Currently the United States imports most of its oil and its dependence on OPEC is growing. Since oil is such a strategic material, how can that vulnerability be addressed? The 2004 U. S. presidential campaign outlined two very different approaches.

President George W. Bush articulated a strategy of increasing domestic production, not only of oil, but also of natural gas and coal. His vision included opening up a portion of the Arctic National Wildlife Refuge for oil drilling. Tax incentives and subsidies would be used to promote domestic production.

Senator John Kerry, on the other hand, sought to promote a much larger role for energy efficiency and energy conservation. Pointing out that expanded domestic production could exacerbate environmental problems (including climate change), he supported such strategies as mandating standards for fuel economy in automobiles and energy efficiency standards in appliances. He was strongly opposed to drilling in the Arctic National Wildlife Refuge.

Over the long run, both candidates favored a transition to a greater reliance on hydrogen as an alternative fuel. Although hydrogen is a clean-burning fuel, its creation can have important environmental impacts; some hydrogen-producing processes (such as those based on coal) pollute much more than others (such as when the hydrogen is created using solar power).

Using economic analysis, figure out what the effects of the Bush and Kerry strategies would be on (1) oil prices in the short run and the long run, (2) emissions affecting climate change, and (3) U.S. imports in the short run and the long run. If you were in charge of OPEC, which strategy would you like to see chosen by Americans? Why?

Why, you might ask, is self-sufficiency so inefficient when embargoes obviously impose so much damage and self-sufficiency could grant immunity from this damage? Why would we want any imports at all when national security is at stake?

The simple answer is that the vulnerability premium is lower than the cost of becoming self-sufficient, but that response merely begs the question, "why is the vulnerability premium lower?" It is lower for three primary reasons: (1) embargoes are not certain events—they may never occur; (2) domestic steps can be taken to reduce vulnerability of the remaining imports; and (3) accelerating domestic production would incur a user cost by lowering the domestic amounts available to future users.

The expected damage caused by one or more embargoes depends on the likelihood of occurrence, as well as the intensity and duration. This means that the P_{w2} curve will be lower for imports having a lower likelihood of being embargoed. Imports from countries less hostile to our interests are more secure and the vulnerability premium on those imports is smaller.[4]

[4]It is this fact that explains the tremendous U.S. interest in Mexican oil, in spite of the fact that, historically, it has not been cheaper.

For any remaining vulnerable imports, we can adopt certain contingency programs to reduce the damage an embargo would cause. The most obvious measure is to develop a domestic stockpile of oil to be used during an embargo. The United States has taken this route. The stockpile, called the *strategic petroleum reserve*, was originally designed to contain one billion barrels of oil (see Example 7.2). A one billion barrel stockpile would replace three million barrels a day for slightly less than one year or a larger number of barrels per day for a shorter period of time. This reserve would serve as an alternative domestic source of supply, which, unlike other oil resources, could be rapidly deployed on short notice. It is, in short, a form of insurance. If this protection can be purchased cheaply, implying a lower P_{w2}, imports become more attractive.

To understand the third and final reason that paying the vulnerability premium would be less costly than self-sufficiency, we must consider vulnerability in a dynamic, rather than static, framework. Because oil is a depletable resource, a user cost is associated with its efficient use. To reorient the extraction of that resource toward the present, as a self-sufficiency strategy would do, reduces future net benefits. Thus, the self-sufficiency strategy tends to be myopic in that it solves the short-term vulnerability problem by creating a more serious one in the future. Paying the vulnerability premium creates a more efficient balance between the present and future, as well as between current imports and domestic production.

We have established the fact that government can reduce our vulnerability to imports, which tends to keep the risk premium as low as possible. Certainly for oil, however, even after the stockpile has been established, the risk premium is not zero; P_{w1} *and* P_{w2} will not coincide. Consequently, the government must also concern itself with achieving both the efficient level of consumption and the efficient share of that consumption borne by imports. Let's examine some of the policy choices.

As noted in Debate 7.1, energy conservation is one popular approach to the problem. One way to accomplish additional conservation is by means of a tax on fossil fuel consumption. Graphically, this approach would be reflected as a shift inward of the after-tax demand curve. Such a tax would reduce energy consumption and emissions of greenhouse gases (an efficient result) but would not achieve the efficient share of imports (an inefficient result). An energy tax falls on *all* energy consumption, whereas the security problem involves only imports. While energy conservation may increase the net benefit, it cannot ever be the sole policy instrument used or an efficient allocation will not be attained.

Another possible strategy employs the subsidization of domestic supply. Diagrammatically, this would be portrayed in Figure 7.4 as a shift of the domestic supply curve to the right. Notice that the effect would be to reduce the share of imports in total consumption (an efficient result) but reduce neither consumption nor climate change emissions (an inefficient result). This strategy also tends to drain domestic reserves faster, which makes the nation more vulnerable in the long run (another inefficient result).

While subsidies of domestic fossil fuels can reduce imports, they will tend to intensify the climate change problem and increase long-run vulnerability. In 2010, the International Energy Agency released *The World Energy Outlook 2010*, a report

EXAMPLE 7.2

Strategic Petroleum Reserve

The U.S. strategic petroleum reserve (SPR) is the world's largest supply of emergency crude oil. The federally owned oil stocks are stored in huge underground salt caverns along the coastline of the Gulf of Mexico.

Decisions to withdraw crude oil from the SPR are made by the President under the authority of the Energy Policy and Conservation Act. In the event of an "energy emergency," SPR oil would be distributed by competitive sale. In practice what constitutes an energy emergency goes well beyond embargoes. The SPR has been used only three times and no drawdown involved protecting against an embargo.

- During Operation Desert Storm in 1991 sales of 17.3 million barrels were used to stabilize the oil market in the face of supply disruptions arising from the war.
- After Hurricane Katrina caused massive damage to the oil production facilities, terminals, pipelines, and refineries along the Gulf regions of Mississippi and Louisiana in 2005, sales of 11 million barrels were used to offset the domestic shortfall.
- A series of emergency exchanges conducted after Hurricane Gustav, followed shortly thereafter by Hurricane Ike, reduced the level by 5.4 million barrels.

The Strategic Petroleum Reserve has never reached the original one billion barrel target, but the Energy Policy Act of 2005 directed the Secretary of Energy to bring the reserve to its authorized one billion barrel capacity. Acquiring the oil to build up the reserve is financed by the Royalty-in-Kind program. Under the Royalty-in-Kind program, producers who operate leases on the federally owned Outer Continental Shelf are required to provide from 12.5 to 16.7 percent of the oil they produce to the U.S. government. This oil is either added directly to the stockpile or sold to provide the necessary revenue to purchase oil to add to the stockpile.

Source: U.S. Department of Energy Strategic Petroleum Reserve Web site: http://www.fe.doe.gov/programs/reserves/index.html and http://www.spr.doe.gov/dir/dir.html (accessed November 1, 2010).

urging nations to eliminate fossil fuel subsidies to curb energy demand and cut the carbon dioxide (CO_2) emissions that cause climate change. They estimated that eliminating fossil fuel subsidies would reduce CO_2 emissions 5.8 percent by 2020. Fossil fuel subsidies were estimated at $312 billion in 2009, compared with $57 billion for renewable energy.

A third approach would tailor the response more closely to the national security problem. One could use either a tariff on imports equal to the vertical distance between P_{w1} *and* P_{w2} or a quota on imports equal to *C*–*B*. With either of these approaches, the price to consumers would rise to P_1, total consumption would fall to *C*, and imports would be *C*–*B*. This achieves the appropriate balance between imports and domestic production (an efficient result), but it does not internalize the climate change cost from using domestic production (an inefficient result).

The use of tariffs or quotas also has some redistributive consequences. Suppose a tariff were imposed on imports equal to the difference between P_{w1} and P_{w2}. The rectangle represented by that price differential times the quantity of imports would then represent tariff revenue collected by the government.

If a quota system were used instead of a tariff and the import quotas were simply given to importers, that revenue would go to the importers rather than the government. This explains why importers prefer quotas to a tariff system.

The effect of either system on domestic producer surplus should also be noticed. Producers of domestic oil would be better off with a tariff or quota on imported oil than without it. Each raises the cost or reduces the availability of the foreign substitute, which results in higher domestic prices. The higher domestic prices induce producers to produce more, but they also result in higher profits on the oil that would have been produced anyway, echoing the premise that public policies may not only restore efficiency but also tend to redistribute wealth.

The Other Depletable Sources: Unconventional Oil and Gas, Coal, and Nuclear Energy

While the industrialized world currently depends on conventional sources of oil and gas for most of our energy, over the long run, in terms of both climate change and national security issues, the obvious solution involves a transition to domestic renewable sources of energy that do not emit greenhouse gases. What role does that leave for the other depletable resources, namely unconventional oil and gas, coal, and uranium?

Although some observers believe the transition to renewable sources will proceed so rapidly that using these fuels will be unnecessary, many others believe that depletable transition fuels will probably play a significant role. Although other contenders do exist, the fuels receiving the most attention (and controversy) as transition fuels are unconventional sources of oil and gas, coal, and uranium. Coal, in particular, is abundantly available, both globally and domestically, and its use frees nations with coal from dependencies on foreign countries.

Unconventional Oil and Gas Sources

The term *unconventional oil and gas* refers to sources that are typically more difficult and expensive to extract than conventional sources. One unconventional source of both oil and natural gas is shale. The flow rate from shale is sufficiently low that oil or gas production in commercial quantities requires that the rock be fractured in order to extract the gas. While gas has been produced for years from shales with natural fractures, the shale gas boom in recent years has been due to a process known as "hydraulic fracturing" (or popularly as "fracking"). To overcome the problem of impermeability, wells

EXAMPLE 7.3

Fuel from Shale: The Bakken Formation

According to the U.S. Geological Service, one of the larger domestic discoveries in recent years of unconventional oil and associated gas can be found in the Bakken Formation in Montana and North Dakota. Parts of the formation extend into the Canadian Provinces of Saskatchewan and Manitoba.

A U.S. Geological Survey assessment, released April 10, 2008, shows some 3–4.3 billion barrels of "technically recoverable" oil in this Formation. (Technically, recoverable oil resources are defined as those producible using currently available technology and industry practices.) This estimate represented a 25-fold increase in the estimated amount of recovered oil compared to the agency's 1995 estimate.

Whereas traditional oil fields produce from rocks with relatively high porosity and permeability, so oil flows out fairly easily, the Bakken Formation consists of low-porosity and -permeability rock, mostly shale, from which oil flows only with difficulty. To overcome this problem, wells are drilled horizontally, at depth, into the Bakken and fracking is used to increase the permeability.

One of the barriers to extracting these resources involves their environmental impact. The U.S. EPA and Congress have noted that serious concerns have arisen from citizens and their representatives about hydraulic fracturing's potential impact on drinking water, human health, and the environment. Concluding that these issues deserve further study, EPA's Office of Research and Development (ORD) will be conducting a scientific study, expected to be completed in 2012, to investigate the possible relationships between hydraulic fracturing and drinking water. Once that study is completed, the future role for the Bakken Formation will likely become clearer.

Sources: 3 to 4.3 Billion Barrels of Technically Recoverable Oil Assessed in North Dakota and Montana's Bakken Formation—25 Times More Than 1995 Estimate—at http://www.usgs.gov/newsroom/article.asp?ID=1911; Hydraulic Fracturing at http://water.epa.gov/type/groundwater/uic/class2/hydraulicfracturing/index.cfm

are drilled horizontally, at depth, and then water and other materials (like sand) are pumped into the well at high pressure to create open fractures, which increase the permeability in these tight rocks. The oil can then flow more easily out of these fractures and tight pores. As Example 7.3 demonstrates, while some of these resources are quite large, they may also pose some difficult environmental and human health challenges.

While many other unconventional oil resources may still be economically out of reach at the present time, two unconventional oil sources are currently being tapped—extra-heavy oil from Venezuela's Orinoco oil belt and bitumen, a tar-like hydrocarbon that is abundant in Canada's tar sands. The Canadian source is particularly important from the U.S. national security perspective, coming as it does, from a friendly neighbor to the North.

The main concern about these sources is also their environmental impact. It not only takes much more energy to extract these unconventional resources (making the net energy gain smaller), but, in the case of Canadian tar sands, large amounts of water are also necessary either to separate bitumen from the sand and other solids, or to produce steam, depending on the oil-recovery method. Emissions of air pollutants, including CO_2, are usually even greater for unconventional sources than they are for conventional sources.

Coal

Coal's main drawback is its contribution to air pollution. High sulfur coal is potentially a large source of sulfur dioxide emissions, one of the chief culprits in the acid-rain problem. It is also a major source of particulate emissions and mercury as well as carbon dioxide, one of the greenhouse gases.

Coal is heavily used in electricity generation and the rate of increase in coal use for this purpose is especially high in China. With respect to climate change, the biggest issue for coal is whether it could be used without adding considerably to greenhouse gas emissions. As the fossil fuel with the highest carbon emissions per unit of energy supplied, that is a tall order.

Capturing CO_2 emissions from coal-fired plants before they are released into the environment and sequestering the CO_2 in underground geologic formations is now technologically feasible. Energy companies have extensive experience in injecting captured carbon dioxide into oil fields as one means to increase the pressure and, hence, increase the recovery rate from those fields. Whether this practice can be extended to saline aquifers and other geologic formations without leakage at reasonable cost is the subject of considerable current research.

Implementing these carbon capture and storage systems require modifications to existing power plant technologies, modifications that are quite expensive. In the absence of any policy controls on carbon emissions, the cost of these sequestration approaches would rule them out simply because the economic damages imposed by failing to control the gases are externalities. The existence of suitable technologies is not sufficient if the underlying economic forces prevent them from being adopted.

Uranium

Another potential transition fuel, uranium, used in nuclear electrical-generation stations, has its own limitations—abundance and safety. With respect to abundance, technology plays an important part. Resource availability is a problem with uranium as long as we depend on conventional reactors. However, if countries move to a new generation of breeder reactors, which can use a wider range of fuel, availability would cease to be an important issue. In the United States, for example, on a heat-equivalent basis, domestic uranium resources are 4.2 times as great as domestic oil and gas resources if they are used in conventional reactors. With breeder reactors, the U.S. uranium base is 252 times the size of its oil and gas base.

With respect to safety, two sources of concern stand out: (1) nuclear accidents, and (2) the storage of radioactive waste. Is the market able to make efficient decisions about the role of nuclear power in the energy mix? In both cases, the answer is no, given the current decision-making environment. Let's consider these issues one by one.

The production of electricity by nuclear reactors involves radioactive elements. If these elements escape into the atmosphere and come in contact with humans in sufficient concentrations, they can induce birth defects, cancer, or death. Although some radioactive elements may also escape during the normal operation of a plant, the greatest risk of nuclear power is posed by the threat of nuclear accidents.

As the accident in Fukushima, Japan in 2011 made clear, nuclear accidents could inject large doses of radioactivity into the environment. The most dangerous of these possibilities is the core meltdown. Unlike other types of electrical generation, nuclear processes continue to generate heat long after the reactor is turned off. This means that the nuclear fuel must be continuously cooled, or the heat levels will escalate beyond the design capacity of the reactor shield. If, in this case, the reactor vessel fractures, clouds of radioactive gases and particulates would be released into the atmosphere.

For some time, conventional wisdom had held that nuclear accidents involving a core meltdown were a remote possibility. On April 25, 1986, however, a serious core meltdown occurred at the Chernobyl nuclear plant in the Ukraine. Although safety standards are generally conceded to be much higher in other industrialized countries than in the countries of the former Soviet Union, the Fukushima accident demonstrated that even higher standards are no guarantee of accident-free operation.

An additional concern relates to storing nuclear wastes. The waste-storage issue relates to both ends of the nuclear fuel cycle—the disposal of uranium tailings from the mining process and spent fuel from the reactors—although the latter receives most of the publicity. Uranium tailings contain several elements, the most prominent being thorium-230, which decays with a half-life of 78,000 years to a radioactive, chemically inert gas, radon-222. Once formed, this gas has a very short half-life (38 days).

The spent fuel from nuclear reactors contains a variety of radioactive elements with quite different half-lives. In the first few centuries, the dominant contributors to radioactivity are fission products, principally strontium-90 and cesium-137. After approximately 1,000 years, most of these elements will have decayed, leaving the transuranic elements, which have substantially longer half-lives. These remaining elements would remain a risk for up to 240,000 years. Thus, decisions made today affect not only the level of risk borne by the current generation—in the form of nuclear accidents—but also the level of risk borne by a host of succeeding generations (due to the longevity of radioactive risk from the disposal of spent fuel).

Nuclear power has also been beset by economic challenges. New nuclear power plant construction became much more expensive, in part due to the increasing regulatory requirements designed to provide a safer system. Its economic advantage over coal dissipated and the demand for new nuclear plants declined. For example, in 1973, in the United States, 219 nuclear power plants were either planned or in operation. By the end of 1998, that number had fallen to 104, the difference being due primarily to cancellations. More recently, after a period with no new applications, high oil prices, government subsidies, and concern over greenhouse gases had caused some resurgence of interest in nuclear power prior to the Fukushima accident.

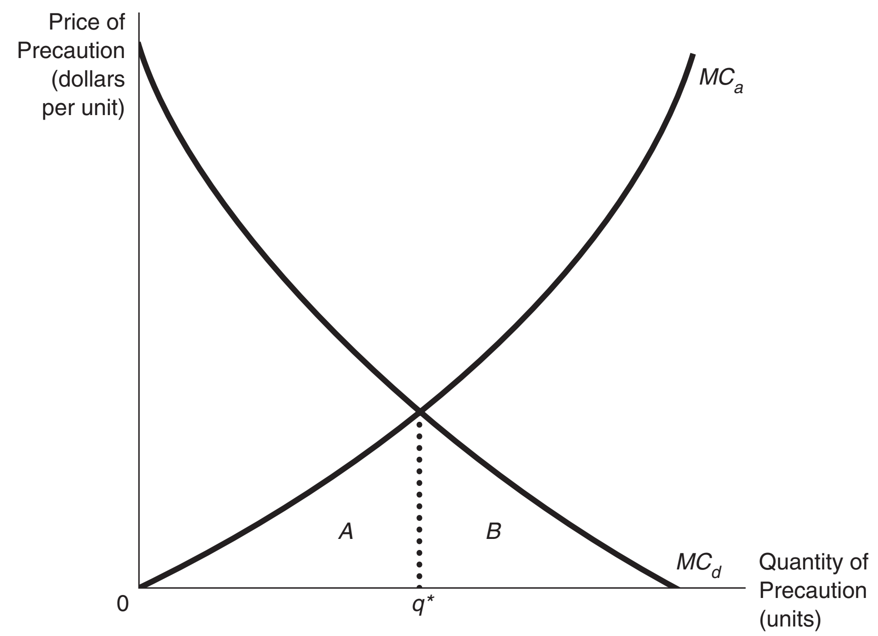
Not all nations have made the same choice with respect to the nuclear option. Japan, along with France, for example, has used standardized plant design and regulatory stability to lower electricity generating costs for nuclear power to the point that they are lower than for coal. Attracted by these lower costs, both countries have relied heavily on nuclear power.

With respect to waste, France chose a closed fuel cycle at the very beginning of its nuclear program, involving reprocessing used fuel so as to recover uranium and plutonium for reuse and to reduce the volume of high-level wastes for disposal. Recycling allows 30 percent more energy to be extracted from the original uranium and leads to a great reduction in the amount of wastes for disposal. What role that nuclear power will play in the future energy plans of these two countries after Fukushima remains to be seen.

Can we expect the market to make the correct choice with respect to nuclear power? We might expect the answer for the problem of nuclear accidents to be no, because this seems to be such a clear case of externalities. Third parties, those living near the reactor, would receive the brunt of the damage from a nuclear accident. Would the utility have an incentive to choose the efficient level of precaution?

If the utility had to compensate fully for all the damages caused, then the answer would be yes. To see why, consider Figure 7.5. Curve MC_a is the marginal cost of damage avoidance. The more precaution that is taken, the higher is the marginal cost. Curve MC_d is the marginal cost of expected damage, suggesting that as more precautionary measures are taken, the additional reduction in damages obtained from those measures declines.

FIGURE 7.5 The Efficient Level of Precaution



The efficient level of precaution is the one minimizing the sum of the costs of precaution and the expected costs of the unabated damage. In Figure 7.5, that point is q^*, and the total cost to society from that choice is the sum of area A and area B.

Will a private utility choose q^*? Presumably it would, if the curves it actually faces are MC_a and MC_d. The utility would be responsible for the costs of precautionary behavior, so it would face MC_a. How about MC_d? We might guess that the utility would face MC_d because people incurring damages could, through the judicial system, sue for damages. In the United States, that guess is not correct for two reasons: (1) the role of the government in sharing the risk, and (2) the role of insurance.

When the government first allowed private industry to use atomic power to generate electricity, there were no takers. No utility could afford the damages if an accident occurred. No insurance company would underwrite the risk. Then in 1957, with the passage of the Price-Anderson Act, the government underwrote the liability. That act provided for a liability ceiling of $560 million (once that amount had been paid, no more claims would be honored), of which the government would bear $500 million. The industry would pick up the remaining $60 million. The act was originally designed to expire in 10 years, at which time the industry would assume full responsibility for the liability.

The act didn't expire, although over time a steady diminution of the government's share of the liability has occurred. Currently the liability ceiling still exists, albeit at a higher level; the amount of private insurance has increased; and a system has been set up to assess all utilities a retrospective premium in the event of an accident.

The effect of the Price-Anderson Act is to shift inward the private marginal damage curve that any utility faces. Both the liability ceiling and the portion of the liability borne by government reduce the potential compensation the utility would have to pay. As the industry assumes an increasing portion of the liability burden and the individual utility assumes less, the risk sharing embodied in the retrospective premium system (the means by which it assumes that burden) breaks the link between precautionary behavior by the individual utility and the compensation it might have to pay. Under this system, increased safety by the utility does not reduce its retrospective premiums.

The cost to all utilities, whether they have accidents or not, is the premium paid both before and after any accident. These premiums do not reflect the amount of precautionary measures taken by an individual plant; therefore, individual utilities have little incentive to provide an efficient amount of safety. In recognition of the utilities' lower-than-efficient concern for safety, the U.S. government has established the Nuclear Regulatory Commission to oversee the safety of nuclear reactors, among its other responsibilities.

To further complicate the problem, the private sector is not the only source of excessive nuclear waste. The U.S. Department of Energy, for example, presides over a nuclear weapons complex containing 15 major facilities and a dozen or so smaller ones.

Both the operating safety and the nuclear waste storage issue can be viewed as a problem of determining appropriate compensation. Those who gain from nuclear power should be forced to compensate those who lose. If they can't, in the absence of externalities, the net benefits from adopting nuclear power are not positive.

If nuclear power is efficient, by definition the gains to the gainers will exceed the losses to the losers. Nonetheless, it is important that this compensation actually be paid because without compensation, the losers can block the efficient allocation.

A compensation approach is already being taken in those countries still expanding the role of nuclear power. The French government, for example, has announced a policy of reducing electricity rates by roughly 15 percent for those living near nuclear stations. And in Japan during 1980, the Tohoku Electric Power Company paid the equivalent of $4.3 million to residents of Ojika, in northern Japan, to entice them to withdraw their opposition to building a nuclear power plant there. Do you think the effectiveness of this approach would be affected by the Fukushima accident?

To the extent it works, this approach could also help resolve the current political controversy over the location of nuclear waste disposal sites. Under a compensation scheme, those consuming nuclear power would be taxed to compensate those who live in the areas of the disposal site. If the compensation is adequate to induce them to accept the site, then nuclear power is a viable option and the costs of disposal are ultimately borne by the consumers. Attracted by the potential for compensation, some towns, such as Naurita, Colorado, have historically sought to become disposal sites.

Are future generations adequately represented in this transaction? The quick answer is no, but that answer is not correct. Those living around the sites will experience declines in the market value of land, reflecting the increased risk of living or working there. The payment system is designed to compensate those who experience the reduction, the current generation. Future generations, should they decide to live near a disposal site, would be compensated by lower land values. If the land values were not cheap enough to compensate them for the risk of that location, they would not have to live there. As long as full information on the risks posed is available, those who do bear the cost of locating near the sites do so only if they are willing to accept the risk in return for lower land values.

Electricity

For a number of electric utilities, conservation has assumed an increasing role. To a major extent, conservation has already been stimulated by market forces. High oil and natural gas prices, coupled with the rapidly increasing cost of both nuclear and coal-fired generating stations, have reduced electrical demand significantly. Yet many regulatory authorities are coming to the conclusion that more conservation is needed.

Perhaps the most significant role for conservation is its ability to defer capacity expansion. Each new electrical generating plant tends to cost more than the last, and frequently the cost increase is substantial. When the new plants come on line, rate increases to finance the new plant are necessary. By reducing the demand for electricity, conservation delays the date when the new capacity is needed. Delays in the need to construct new plants translate into delays in rate increases as well.

Governments are reacting to this situation in a number of ways. One is to promote investments in conservation, rather than in new plants, when conservation is the cheaper alternative. Typical programs have established systems of rebates for residential customers to install conservation measures in their homes, provided free home weatherization to qualified low-income home owners, offered owners of multifamily residential buildings incentives for installing solar water heating systems, and provided subsidized energy audits to inform customers about money-saving conservation opportunities. Similar incentives have been provided to the commercial, agricultural, and industrial sectors.

The total amount of electric energy demanded in a given year is not the only concern. How that energy demand is spread out over the year is also a concern. The capacity of the system must be high enough to satisfy the demand even during the periods when the energy demand is highest (called *peak periods*). During other periods, much of the capacity remains underutilized.

Demand during the peak period imposes two rather special costs on utilities. First, the peaking units, those generating facilities fired up only during the peak periods, produce electricity at a much higher marginal cost than do base-load plants, those fired up virtually all the time. Typically, peaking units are cheaper to build than base-load plants, but they have higher operating costs. Second, it is the growth in peak demand that frequently triggers the need for capacity expansion. Slowing the growth in peak demand can delay the need for new, expensive capacity expansion, and a higher proportion of the power needs can be met by the most efficient generating plants.

Utilities respond to this problem by adopting load-management techniques to produce a more balanced use of this capacity over the year. One economic load-management technique is called *peak-load pricing*. Peak-load pricing attempts to impose the full (higher) marginal cost of supplying peak power on those consuming peak power by charging higher prices during the peak period.

While many utilities have now begun to use simple versions of this approach, some are experimenting with innovative ways of implementing rather refined versions of this system. One system, for example, transmits electricity prices every five minutes over regular power lines. In a customer's household, the lines attached to one or more appliances can be controlled by switches that turn the power off any time the prevailing price exceeds a limit established by the customer. Other less sophisticated pricing systems simply inform consumers, in advance, of the prices that will prevail in predetermined peak periods.

Studies by economists indicate that even the rudimentary versions of peak-load pricing work. The greatest shifts are typically registered by the largest residential customers and those with several electrical appliances.

Also affecting energy choices is the movement to deregulate electricity production. Historically, electricity was generated by regulated monopolies. In return for accepting both government control of prices and an obligation to service all customers, utilities were given the exclusive rights to service-specific geographic areas.

In the 1990s, it was recognized that while electricity distribution has elements of a natural monopoly, generation does not. Therefore, several states and a number of national governments have deregulated the generation of electricity, while keeping the distribution under the exclusive control of a monopoly. In the United States, electricity deregulation officially began in 1992 when Congress allowed independent energy companies to sell power on the wholesale electricity market.

Forcing generators to compete for customers, it was believed, would produce lower electricity bills for customers. Experience reveals that lower prices have not always been the result (see Example 7.4).

Electricity deregulation has also raised some environmental concerns. Since electricity costs typically do not include all the costs of environmental damage, the sources offering the lowest prices could well be highly polluting sources. In this case, environmentally benign generation sources would not face a level playing field; polluting sources would have an inefficient advantage.

One policy approach for dealing with these concerns involves renewable energy credits (RECs). Renewable energy sources, such as wind or solar, are frequently characterized by relatively large capital costs, relatively low variable costs (since the fuel is costless), and low pollution emissions. Energy markets may ignore the advantages of low pollution emissions (since pollution imposes an external cost) and are likely to be characterized by short-term energy sales and price volatility (to the detriment of investors, who usually prefer investments with low capital costs and short payback periods). Under these circumstances, investments in capital-intensive, renewable energy technologies are unlikely to be sufficient to achieve an efficient outcome.

Renewable energy credits are designed to facilitate the transition to renewable power by overcoming these obstacles. A generator of electricity from a renewable source (such as wind or photovoltaics) can produce two saleable commodities. The first is the electricity itself, which can be sold to the grid, while the second is the renewable energy credit that turns the environmental attributes (such as the fact that it was created by a qualifying renewable source) into a legally recognized form of property that can be sold separately. Generally renewable generators create one REC for every 1,000 kilowatt-hours (or, equivalently, 1 megawatt-hour) of electricity placed on the grid.

The demand for these credits comes from diverse sources, but the most prominent are: (1) voluntary markets, involving consumers or institutions that altruistically choose to support green electricity and (2) compliance markets, involving electricity generators that need to comply with a renewable energy standard.

Some states with restructured electricity markets authorize voluntary markets in which households or institutions can directly buy green power, if offered (typically at a higher price) by their generator, or by purchasing RECs if their current provider does not offer green power. This allows consumers or institutions to lower their own carbon footprint since the REC they purchase and retire represents a specific amount of avoided greenhouse gas emissions. Educational institutions, for example, are incorporating the purchase of RECs into their strategies for achieving the goal of carbon neutrality adopted after signing onto the American College & University Presidents' Climate Commitment. As of August 2010, some 30 REC retail products were available to consumers and institutions.

The compliance market, apparently the larger of the two, has arisen because some states have imposed renewable portfolio standards (RPS) on electricity generators. Requiring a certain percent of electricity in the jurisdiction be generated from qualified renewable power sources, these standards can either be met by actually generating the electricity from qualified sources or purchasing a sufficient number of RECs from generators that have produced a higher percent

EXAMPLE 7.4

Electricity Deregulation in California: What Happened?

In 1995, the state legislature of California reacted to electricity rates that were 50 percent higher than the U.S. average by unanimously passing a bill to deregulate electricity generation within the state. The bill had three important features: (1) all utilities would have to divest themselves of their generation assets; (2) retail prices of electricity would be capped until the assets were divested; and (3) the utilities were forced to buy power in a huge open-auction market for electricity, known as a spot market, where supply and demand were matched every day and hour.

The system was seriously strained by a series of events that restricted supplies and raised prices. Although the fact that demand had been growing rapidly, no new generating facility had been built over a decade and much of the existing capacity was shut down for maintenance. An unusually dry summer reduced generating capacity at hydroelectric dams and electricity generators in Oregon and Washington, traditional sources of imported electricity. In addition, prices rose for the existing supplies of natural gas, a fuel that supplied almost one-third of the state's electricity.

This combination of events gave rise to higher wholesale prices, as would be expected, but the price cap prevented them from being passed on to consumers. Since prices could not equilibrate the retail market, blackouts (involving a complete loss of electricity to certain areas at certain times) resulted. To make matters worse, the evidence suggests that wholesale suppliers were able to take advantage of the short-term inflexibility of supply and demand to withhold some power from the market, thereby raising prices more and creating some monopoly profits. And on April 6, 2001, Pacific Gas and Electric, a utility that served a bit more than one-third of all Californians, declared bankruptcy.

Why had a rather simple quest for lower prices resulted in such a tragic outcome? Are the deregulation plans in other states headed for a similarly dismal future? Time will tell, of course, but that outcome seems unlikely. A reduction of supplies could affect other areas, though the magnitude of the confluence of events in California seems unusually harsh. Furthermore, the design of the California deregulation plan was clearly flawed. The price cap, coupled with the total dependence on the spot market, created a circumstance in which the market not only could not respond to the shortage but in some ways made it worse. Since neither of those features is an essential ingredient of a deregulation plan, other areas can choose more prudent designs.

Sources: Severin Borenstein, Jim Bushnell, and Frank Wolak. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market," A paper presented at the American Association meetings in Atlanta, January 2001; P. L. Joskow. "California's Electricity Market Meltdown," *Economies et Sociétés* Vol. 35, No. 1–2 (January–February 2001): pp. 281–296.

from those sources than the mandate. By providing this form of flexibility in how the mandate is met, RECs lower the compliance cost, not only in the short run (by allowing the RECs to flow to the areas of highest need), but also in the long run (by making renewable source generation more profitable in areas not under a renewable energy mandate than it would otherwise be).

Although by 2010 some 38 states and the District of Columbia had a renewable energy standard and a majority of those have REC programs, RECs are no panacea. Experience in several U.S. states shows that a poorly designed system does little to increase renewable generation (Rader, 2000). On the other hand, appropriately designed systems can provide a significant boost to renewable energy (see Example 7.5). The details matter.

Another innovation in the electric power industry has also given rise to a new market trading a new commodity.[5] Known as the forward capacity market, this approach uses market forces to facilitate the planning of future electric capacity investment.

The Independent System Operator for New England (ISO-NE) is the organization responsible for ensuring the constant availability of electricity, currently and for future generations, in the New England area. ISO-NE meets this obligation in three ways: by ensuring the day-to-day reliable operation of New England's bulk power generation and transmission system, by overseeing and administering the region's wholesale electricity markets, and by engaging in comprehensive, regional planning processes.

EXAMPLE 7.5

Tradable Energy Credits: The Texas Experience

Texas has rapidly emerged as one of the leading wind power markets in the United States, in no small part due to a well-designed and carefully implemented renewable portfolio standard (RPS coupled with renewable energy credits. While the RPS specifies targets and deadlines for producing specific proportions of electricity from renewable resources (wind, in this case), the credits lower compliance cost by increasing the options available to any party required to comply.

The early results have been impressive. Initial RPS targets in Texas were easily exceeded by the end of 2001, with 915 megawatts of wind capacity installed in that year alone. The response has been sufficiently strong that it has become evident that the RPS capacity targets for the next few years would be met early. RPS compliance costs are reportedly very low, in part due to a complementary production tax credit (a subsidy to the producer), the especially favorable wind conditions in Texas, and an RPS target that was ambitious enough to allow economies of scale to be exploited. The fact that the cost of administering the program is also low, due to an efficient Web-based reporting and accounting system, also helps.

Finally, and significantly, retail suppliers have been willing to enter into long-term contracts with renewable generators, reducing exposure of both producers and consumers to potential volatility of prices and sales. Long-term contracts ensure developers a stable revenue stream and, as a result, access to low-cost financing, while offering customers a reliable, steady supply of electricity.

Sources: O. Langniss and R. Wiser. "The Renewables Portfolio Standard in Texas: An Early Assessment," *Energy Policy* Vol. 31 (2003): pp. 527–535; N. Rader. "The Hazards of Implementing Renewable Portfolio Standards," *Energy and Environment,* Vol. 11, No. 4 (2000): pp. 391–405; L. Nielsen and T. Jeppesen. "Tradable Green Certificates in Selected European Countries—Overview and Assessment," *Energy Policy* Vol. 31 (2003): pp. 3–14; and The Texas Renewable Credit website http://www.texasrenewables.com/

[5] For details on this market, see http://www.iso-ne.com/markets/othrmkts_data/fcm/index.html.

The objective of the Forward Capacity Market (FCM), run by ISO-NE, is to assure that sufficient peak generating capacity for reliable system operation for a future year will be available. Since ISO-NE does not itself generate electricity, to assure this future capacity, they solicit bids in a competitive auction not only for additional generating capacity, but also for legally enforceable additional reductions in peak demand from energy efficiency, which reduces the need for more capacity. This system allows strategies for reducing peak demand to complete on a level playing field with strategies to expand capacity.

Another quite different approach to promoting the use of renewable resources in the generation of electric power is known as a *feed-in tariff*. Used in Germany, this approach focuses on establishing a stable price guarantee rather than a subsidy or a government mandate (see Example 7.6).

EXAMPLE 7.6

Feed-in Tariffs

Promoting the use of renewable resources in the generation of electricity is both important and difficult. Germany provides a very useful example of a country that seems to be especially adept at overcoming these barriers. According to one benchmark, at the end of 2007, renewable energies were supplying more than 14 percent of the electricity used in Germany, exceeding the original 2010 goal of 12.5 percent.

What prompted this increase? The German feed-in tariff determines the prices received by anyone who installs qualified renewable capacity producing electricity for the grid. In general, a fixed incentive payment per kilowatt-hour is guaranteed for that installation. The level of this payment (determined in advance by the rules of the program) is based upon the costs of supplying the power and is set at a sufficiently high level so as to assure installers that they will receive a reasonable rate of return on their investment. While this incentive payment is guaranteed for 20 years for each installed facility, each year the level of that guaranteed 20-year payment is reduced (typically in the neighborhood of 1–2 percent per year) for new facilities to reflect expected technological improvements and economies of scale.

This approach has a number of interesting characteristics:

- It seems to work.
- No subsidy from the government is involved; the costs are borne by the consumers of the electricity.
- The relative cost of the electricity from feed-in tariff sources is typically higher in the earlier years than for conventional sources, but lower in subsequent years (as fossil fuels become more expensive). In Germany the year in which electricity becomes cheaper due to the feed-in tariff is estimated to be 2025.
- This approach actually offers two different incentives: (1) it provides a price high enough to promote the desired investment and (2) it guarantees the stability of that price rather than forcing investors to face the market uncertainties associated with fluctuating fossil fuel prices or subsidies that come and go.

Source: Jeffrey H. Michel. (2007). "The Case for Renewable Feed-In Tariffs." Online Journal of the EUEC, Volume 1, Paper 1, available at http://www.euec.com/journal/Journal.htm

Energy Efficiency

As the world grapples with creating the right energy portfolio for the future, energy efficiency policy is playing an increasingly prominent role. In recent years the amount of both private and public money being dedicated to promoting energy efficiency has increased a great deal.

The role for energy efficiency in the broader mix of energy polices depends, of course, on how large the opportunity is. Estimating the remaining potential is not a precise science, but the conclusion that significant opportunities remain seems inescapable.

The existence of these opportunities can be thought of as a necessary, but not sufficient condition for government intervention. Depending upon the level of energy prices and the discount rate, the economic return on these investments may be too low to justify intervention. Additionally, policy intervention could, in principle, be so administratively costly as to outweigh any gains that would result.

The strongest case for government intervention flows from the existence of externalities. Markets are not likely to internalize these external costs on their own. The natural security and climate change externalities mentioned above, as well as other external co-benefits such as pollution-induced community health effects, certainly imply that the market undervalues investments in energy efficiency.

The analysis provided by economic research in this area, however, makes it clear that the case for policy intervention extends well beyond externalities. Internalizing externalities is a very important, but incomplete, policy response.

Consider just a few of the other foundations for policy intervention. Inadequately informed consumers can impede rational choice, as can a limited availability of capital (preventing paying more up front for the more energy-efficient choice even when the resulting energy savings would justify the additional expense in present value terms). Perverse incentives can also play a role as in the case of one who lives in a room (think dorm) or apartment where the amount of energy used is not billed directly, resulting in a marginal cost of additional energy use that is zero.

A rather large suite of policy options has been implemented to counteract these other sources of deficient levels of investment in energy efficiency. Some illustrations include the following:

- Certification programs such as Energy Star for appliances or LEED (Leadership in Energy and Environmental Design) standards for buildings attempt to provide credible information for consumers to make informed choices on energy efficiency options.
- Minimum efficiency standards (e.g., for appliances) prohibit the manufacture, sale, or importation of clearly inefficient appliances.
- An increased flow of public funds into the market for energy efficiency has led to an increase in the use of targeted investment subsidies. The most common historic source of funding in the electricity sector involved the use of a small mandatory per kilowatt-hour charge (typically called a “system benefit charge” or “public benefit charge”) attached to the distribution service bill. The newest source of funding comes from the revenue accrued

from the sale of carbon allowances in several state or regional carbon cap-and-trade programs (described in detail in Chapter 14). The services funded by these sources include supplementing private funds for diverse projects such as weatherization of residences for low-income customers to more efficient lighting for commercial and industrial enterprises.

The evidence suggests that none of these policies either by themselves or in concert are completely efficient, but that they have collectively represented a move toward a more efficient use of energy. Not only does the evidence seem to suggest that they have been effective in reducing wasteful energy demand, but also that the programs have been quite cost-effective, with program costs well below the cost of the alternative, namely generating the energy to satisfy that demand.

Transitioning to Renewables

Ultimately our energy needs will have to be fulfilled from renewable energy sources, either because the depletable energy sources have been exhausted or, as is more likely, the environmental costs of using the depletable sources have become so high that renewable sources will be cheaper.

One compelling case for the transition is made by the mounting evidence that the global climate is being jeopardized by current and prospective energy consumption patterns. (A detailed analysis of this problem is presented in Chapter 16.) To the extent that rapidly developing countries, such as China and India, were to follow the energy-intensive, fossil fuel–based path to development pioneered by the industrialized nations, the amount of CO_2 released into the air would be unprecedented. A transition away from fossil fuels to other energy forms in both the industrialized and developing nations would be an important component in any strategy to reduce CO_2 emissions. Can our institutions manage that transition in a timely and effective manner?

Renewable energy comes in many different forms. It is unlikely that any one source will provide the long-run solution, in part because both the timing (peak demands) and form (gases, liquids, or electricity) of energy matter. Different sources will have different comparative advantages so, ultimately, a mix of sources will be necessary. Consider some of the options. The extent to which these sources will penetrate the market will depend upon their relative cost and consumer acceptance.

Hydroelectric Power

Hydroelectric power, which is generated when turbines convert the kinetic energy from a flowing body of water into electricity, passed that economic test a long time ago and is already an important source of power. This source of power is clean from an emissions point of view and domestic hydropower can help with national security concerns as well. On the other hand, hydroelectric dams can be a significant impediment to fish migrations and water quality. The impounded water can flood ecosystems and displace villages, and the buildup of silt behind the dams not only can lower the life of the facility, but also can alter the upstream and downstream ecosystems.

According to the U.S. Department of Energy (DOE),[6] hydropower in the United States rose from 15 percent of electricity generation in 1907 to 40 percent in 1940, but fell to only 10 percent by 2003. DOE estimates that some 80,000 MW of hydro power are currently operating in the United States and their resource assessment identified 5,677 sites in the United States with an undeveloped capacity of about 30,000 MW.

While hydroelectric power has been cost-effective for some time, other renewable resources have not been. Some, like wind, have become cost-effective, while the cost-effectiveness of others awaits further technological developments or additional increases in the costs of fossil fuels. This would occur, for example, by the imposition of a carbon tax or cap-and-trade policy to internalize the climate-change externality.

Wind

Wind power is beginning to penetrate the market on a rather large scale. New designs for turbines that convert wind energy to electricity have reduced the cost and increased the reliability of wind-generated electricity to the point that it now can compete with conventional sources in favorable sites even when environmental costs have not been internalized. (Favorable sites are those with sufficiently steady, strong winds.) Although many unexploited favorable sites still exist around the world, the share of wind power in the total energy mix will ultimately be limited as availability of unexploited sites diminishes.

Wind also has environmental effects that have triggered strong local controversies (see Debate 7.2).

Photovoltaics

Technological change can lower the relative cost of renewable resources. Perhaps the best example of how research can lower costs is provided by the experience with photovoltaics. Photovoltaics involve the direct conversion of solar energy to electricity (as opposed to indirect conversions, such as when solar-heated steam is used to drive a turbine). Anticipating a huge potential market, private industry has been very interested in photovoltaics and has poured a lot of research dollars into improving its commercial viability. The research has paid off. In 1976, the average market price for photovoltaic modules was $30. By 2002 this price had already fallen to $3.75.[7] Rural electrification projects using photovoltaics are slowly spreading into developing countries. Their attractiveness is especially high in regions that have not already established a traditional grid system of large generators and distribution lines. Photovoltaic systems allow these countries to provide electricity to remote regions while avoiding the very high capital cost associated with expanding traditional grid systems into those areas.

---

[6] U.S. Department of Energy Web site: http://www1.eere.energy.gov/windandhydro/hydro_history.html

7U.S. Department of Energy Renewable Energy Annual (REA): http://www.eia.doe.gov/cneaf/solar.renewables/page/rea_data/rea_sum.html

DEBATE 7.2

Dueling Externalities: Should the United States Promote Wind Power?

On the surface the answer seems like a no-brainer, since wind power is a renewable energy source that emits no greenhouse gases, unlike all the fossil fuels it would be likely to replace. Yet some highly visible, committed environmentalists, including Robert F. Kennedy, Jr., have strongly opposed wind projects. Why has this become such a public contentious issue?

Opposition to wind power within the environmental community arises for a variety of reasons. Some point out that the turbines can be noisy for those who live, camp, or hike nearby. Others note that these very large turbines can be quite destructive to bats and birds, particularly if they are constructed in migratory pathways. And a number of opponents object to the way the view would be altered by a large collection of turbines on otherwise-pristine mountaintops or off the coast.

Both the benefits from wind power (reduced impact on the climate) and the costs (effects on aesthetics, birds and noise) are typically externalities. This implies that the developers and consumers of wind power will neither reap all of the environmental benefits from reduced impact on the climate, nor will they typically bear the environmental costs. Making matters even more difficult some of the environmental costs will be concentrated on a relatively few people (those living nearby, for example), while the benefits will be conferred on all global inhabitants, many of whom will bear absolutely no costs whatsoever. The concentrated costs may be an effective motivator to attend the hearings, which are likely to be held near the proposed site, but the diffuse benefits will likely not be.

Since the presence of externalities typically undermines the ability of a market to produce an efficient outcome, it is not surprising that the permitting process for new wind power facilities is highly regulated. Regulatory processes generally encourage public participation by holding hearings. With environmental externalities lying on both sides of the equation and with many of the environmental costs concentrated on a relatively small number of people, it is neither surprising that the hearings have become so contentious, nor that the opposition to wind power is so well represented.

Source: Robert F. Kennedy Jr. "An Ill Wind Off Cape Cod." THE NEW YORK TIMES, Op-Ed, December 16, 2005; and Felicity Barringer, "Debate over Wind Power Creates Environmental Rift." THE NEW YORK TIMES, June 6, 2006.

Active and Passive Solar Energy

The sun's energy can also be used for heating in either an active or passive mode. The difference between the two is that while the passive mode uses no external energy sources, the active mode may use nonsolar energy to power pumps or fans. Solar energy can be used to provide space heating or to provide hot water.

Since the input energy comes from the sun's rays, it is costless; but the system to collect those rays, to transform them into useful heat, and to distribute the heat requires a capital investment. When storage or backup systems are required, they add to the cost.

Ocean Tidal Power[8]

One energy source that relies on the natural cycles of the earth is tidal power. It capitalizes on the fact that coastal areas experience two high and two low tides in a period of time somewhat longer than 24 hours. The energy in the water as it rushes in or out of an inlet or cove is transformed into electricity by a conversion device, commonly a turbine. According to the U.S. Department of Energy, for the tidal differences to be harnessed into electricity, the difference between high and low tides must be at least 5 meters, or more than 16 feet. Only about 40 sites on the earth have tidal ranges of this magnitude. Although no tidal power plants currently are operating in the United States, conditions are good for tidal power generation in both the Pacific Northwest and the Atlantic Northeast regions of the country.

Tidal power plants are not without their environmental impacts. Some designs can impede sea life migration, and silt buildups behind such facilities can impact local ecosystems.

Like many other renewable sources, the input energy is free, but construction costs are high. As a result, the U.S. Department of Energy estimates that the cost per kilowatt-hour of tidal power is currently not competitive with conventional fossil fuel power, but internalizing all the external costs of fossil fuel power could affect that outlook.

Liquid Biofuels

Liquid biofuels, which are made from plant material, are currently receiving a lot of attention in policy circles because potentially they can reduce greenhouse gases and imports of oil simultaneously. These include two alcohols—ethanol and methanol—and biodiesel, an oxygenated fuel produced from a range of biomass-derived feedstocks, including oilseeds, waste vegetable oils, cooking oil, and even animal fats. How cost-effective are they? Hill et al. (2006) examine this issue in detail and provide some useful insights:

- Ethanol yields 25 percent more energy than the energy invested in its production, whereas biodiesel yields 93 percent more.
- Compared with ethanol, biodiesel releases just 1 percent, 8.3 percent, and 13 percent of the agricultural nitrogen, phosphorus, and pesticide pollutants, respectively, per net energy gain.
- Relative to the fossil fuels they displace, greenhouse gas emissions are reduced 12 percent by the production and combustion of ethanol and 41 percent by biodiesel. The advantages of biodiesel over ethanol come from lower agricultural inputs and more efficient conversion of feedstocks to fuel.

How about economic impacts? The authors also point out that neither of these biofuels can replace much petroleum without impacting food supplies and costs,

[8]The information in this section was derived from http://www.eere.energy.gov/consumer/renewable_energy/ocean/index.cfm/mytopic=50008

and those impacts could be serious. We pay more attention to this choice between using crops for food or fuel in Chapter 11.

The bottom line is that both the type of fuel produced and the type of biomass used to produce it matter. Biodiesel has significant energy and environmental benefits over ethanol. Furthermore, biofuels produced from low-input biomass grown on agriculturally marginal land or from waste biomass would, they believe, provide much greater supplies and stronger environmental benefits than food-based biofuels. This analysis certainly raises serious questions about the wisdom of the current U.S. approach that relies heavily on subsidizing ethanol from corn.

Geothermal Energy

A rather different source, geothermal energy is derived from the earth's heat. In some places, geothermal reservoirs of steam or hot water occur where hot magma comes close enough to the surface to heat groundwater to quite high temperatures. When the temperature of the geothermal water reaches 220 degrees Fahrenheit or higher, geothermal energy can be used to generate electricity.

In other places, geothermal can be used even if the water temperatures are much more moderate. When the temperature of a geothermal source is around 50 degrees Fahrenheit and higher, it can be used in combination with heat pumps to provide space heating in the winter and cooling (air-conditioning) in the summer. (Heat pumps are electric devices that use compression and decompression of gas to heat and/or cool a house. Geothermal heat pumps are similar to ordinary heat pumps, but they use the geothermal water resource instead of outside air as the input source for the pump.)

Geothermal systems generally have a higher initial (capital) cost than alternative heating and cooling systems. Based on the estimated yearly energy and maintenance cost savings, the payback period (the number of years it takes for an investor to recover the capital cost from annual cost savings) for a geothermal heat pump system can vary from two to ten years.

Hydrogen

One fuel that is currently receiving intense interest for the long run is hydrogen. (Iceland, for example, has announced its intention to become a hydrogen-fueled economy.) Although hydrogen is the most plentiful element in the universe, it is normally combined with other elements. Water, for example, combines two atoms of hydrogen with one atom of oxygen (H_2O). Hydrogen is also found in "hydrocarbons" that make up many of the fossil fuels, such as gasoline, natural gas, methanol, and propane.

Hydrogen can be made by separating it from hydrocarbons, using heat. Currently, most hydrogen is made this way from natural gas. Alternatively, it can be produced by separating the hydrogen and oxygen atoms in water. If an electric current (produced by photovoltaics, for example) is conducted through a reservoir of water, the liquid splits into its constituent elements. NASA has used liquid hydrogen since the 1970s to propel the space shuttle and other rockets into orbit.

In addition to being directly combustible, hydrogen can be used in fuel cells. Fuel cells offer a promising technology for use as a source of heat and electricity for buildings, and as an electrical power source for electric vehicles. Hydrogen fuel cells powered the NASA space shuttle's electrical systems, producing a clean by-product—pure water—which the crew drank.

Several barriers must be dealt with if the hydrogen-based economy is to become a reality. The technologies that use hydrogen as a fuel are currently very expensive, and the infrastructure needed to get the hydrogen to users is undeveloped.

It is unlikely that hydrogen will be fully competitive with more conventional fuels in the absence of a specific role for government. One potentially substantial cost savings from using hydrogen, the reduction in air pollution damage, is an externality. Since consumers are likely to ignore, or at least weigh less, external costs in their choice of fuels, in the absence of corrective government policy (such as a tax on more polluting fuels), demand will be biased away from hydrogen, and potential suppliers will be discouraged from entering the market.

Consumer acceptance is an important ingredient in the transition to any alternative source of energy. New systems are usually less reliable and more expensive than old systems. Once they become heavily used, reliability normally increases and cost declines; experience is a good teacher. Since the early consumers, the pioneers, experience both lower reliability and higher costs, procrastination can be an optimal individual strategy. Waiting until all the bugs have been worked out and costs come down reduces uncertainty. If every consumer procrastinates about switching, however, the industry will not be able to operate at a sufficient scale and will not be able to gain enough experience to produce the reliability and lower cost that will ensure a large, stable market. How can this initial consumer reluctance be overcome?

One strategy involves using tax dollars to subsidize purchases by the pioneers. Once the market is sufficiently large that it can begin to take advantage of economies of scale and eliminate the initial sources of unreliability, the subsidies could be eliminated. The available empirical evidence based upon the impact of earlier policies (Durham et al., 1988; Fry, 1986) suggests that the tax credit approach did increase the degree of market penetration of solar equipment in the United States.

In the United States, substantial tax credits authorized at both the federal and state levels have been influential in inducing independent producers to accept the financial and engineering risks associated with developing wind power. Although the initial federal tax credits expired in 1985, a 1.5¢/Kwh production incentive for producers of electricity generated from wind power was reinstated in 1992. Since that time the subsidy has elapsed and been reinstated irregularly.

In contrast to the "on-again, off-again" nature of the U.S. subsidies, European nations have been steadily increasing the economic incentives for encouraging wind power, with the result that Europe is now dominating the production of wind power. An alternative approach would involve removing inefficient subsidies or internalizing the externalities for competing fuels in order to create a level playing field for sustainable energy sources. In the absence of those steps, decisions that depend on private, not social, costs will inefficiently favor polluting sources.

Removing subsidies has a certain political appeal. Removal can lower government expenditures (or raise tax revenue in the case of eliminating tax breaks), welcome news during periods of tight budgets. The fact that removal could improve efficiency does not, however, mean that this step is easily taken. The producers of favored energy sources clearly benefit from those subsidies and would fight their removal.

Summary

We have seen that the relationship between government and the market is not always harmonious and efficient. In the past, price controls have tended to reduce energy conservation, to discourage exploration and supply, to cause biases in the substitution among fuel types that penalize future consumers, and to create the potential for abrupt, discontinuous transitions to renewable sources. This important example makes a clear case for less, not more, regulation.

This conclusion is not universally valid, however. Other dimensions of the energy problem, such as climate change and national security issues, suggest the need for some government role. Insecure foreign sources require policies such as tariffs and strategic reserves to reduce vulnerability and to balance the true costs of imported and domestic sources. In addition, government must ensure that the costs of energy fully reflect not only the potentially large environmental costs, including climate change, but also the national security costs associated with our dependence on foreign sources of energy. Government action must also assure that inefficient subsidies do not undermine the transition to sustainable energy resources.

Economic analysis reveals that no single strategy is sufficient to solve the national security and climate change problems simultaneously. Subsidizing domestic supply, for example, would reduce the share of imports in total consumption (an efficient result), but it would reduce neither consumption nor climate change emissions (inefficient results). The expansion of domestic production also tends to drain domestic reserves faster, which makes the nation more vulnerable in the long run (another inefficient result). On the other hand, energy conservation (promoted by a tax on energy, for example) would reduce energy consumption and emissions of greenhouse gases (efficient outcomes) but would not achieve the efficient share of imports (an inefficient result) since an energy tax falls on all energy consumption, whereas the national security problem involves only imports.

Given the inefficient biases against public safety in the Price-Anderson Act, government should also continue to oversee nuclear reactor safety and should ensure that communities accepting nuclear waste disposal sites are fully compensated. Given the environmental difficulties with all three of the depletable transition fuels (unconventional oil, coal, and uranium), energy efficiency, energy conservation, and electric load-management techniques are now playing (and will continue to play) a larger role.

The menu of energy options as the economy transitions to renewable sources offers a large number of choices, including photovoltaics, active and passive solar energy, wind, ocean tidal power, biomass fuels, geothermal energy, and hydrogen. It is far from clear what the ultimate mix will turn out to be, but it is very clear that government policy is a necessary ingredient in any smooth transition to a sustainable-energy future. Since many of the most important costs of energy use are externalities, an efficient transition to these renewable sources will not occur unless the playing field is leveled. The potential for an efficient and sustainable allocation of energy resources by our economic and political institutions clearly exists, even if historically it has not always occurred.

Discussion Questions

1. Should benefit–cost analysis play the dominant role in deciding the proportion of electric energy to be supplied by nuclear power? Why or why not?
2. Economist Abba Lerner proposed a tariff on oil imports equal to 100 percent of the import price. This tariff is designed to reduce dependence on foreign sources as well as to discourage OPEC from raising prices (since, due to the tariff, the delivered price would rise twice as much as the OPEC increase, causing a large subsequent reduction in consumption). Should this proposal become public policy? Why or why not?
3. Does the fact that the strategic petroleum reserve has never been used to offset shortfalls caused by an embargo mean that the money spent in creating the reserve has been wasted? Why or why not?

Self-Test Exercises

1. During a worldwide recession in 1983, the oil cartel began to lower prices. Why would a recession make the cartel more vulnerable to price cutting? How would the reduced demand be shared between the competitive fringe and the cartel members in the absence of this price cutting?
2. Assume the demand and marginal cost conditions given in the second self-test exercise in Chapter 2. In addition, assume that the government imposes a price control at $P = \$80/3$. (a) Find the consumer and producer surplus associated with the resulting allocation. (b) Compare this price control allocation to the monopoly allocation in part (c) of that self-test exercise .
3. Not long ago, a conflict between a paper company and a coalition of environmental groups arose over the potential use of a Maine river for hydroelectric power generation. As one aspect of its case for developing the dam, the paper company argued that without hydroelectric power the energy cost of operating five particular paper machines would be so high that they

would have to be shut down. Environmental groups countered that the energy cost was estimated to be too high by the paper company because it was assigning all of the high-cost (oil-fired) power to these particular machines. That was seen as inappropriate because all machines were connected to the same electrical grid and therefore drew power from all sources, not merely the high-cost sources. They suggested, therefore, that the appropriate cost to assign to the machines was the much lower average cost. Revenue from these machines was expected to be sufficient to cover this average cost. Who was right?

4. In the section of this chapter dealing with load management by the utilities, it was mentioned that peaking plants are typically cheap to build (compared to base-load plants), but that they have relatively high operating costs. Explain why it makes sense for utilities to use this lower-capital, high-operating-cost type of plant for peaking and the high-capital, lower-operating-cost type of plant for base load.
5. If OPEC raised the price of oil high enough, would that be sufficient to promote an efficient energy mix?
6. Label the following as True, False, or Uncertain and explain your choice. (Uncertain means that it can be either true or false depending upon the circumstances.)
   a. All members of a resource cartel share a common objective—increase prices as much and as soon as possible.
   b. By holding prices lower than they would otherwise be, placing a price control on a depletable resource increases both the speed with which the resource is extracted over time and the cumulative amount ultimately extracted.
   c. A price control actually has no influence on the extraction path of a depletable resource until such time as the market price actually reaches the level of the price control.
   d. Forcing companies that drill offshore for oil to compensate victims of any oil spill from one of its facilities would be an efficient requirement.
7. Explain why the existence of a renewable energy credit market would lower the compliance costs for utilities forced to meet a renewable portfolio standard.
8. Using Figure 7.4, show how the level of oil imports and the price level would be affected if the country represented in that figure acted to internalize national security issues, but ignored climate change impacts.

Further Reading

Anthoff, David, and Robert Hahn. "Government Failure and Market Failure: On the Inefficiency of Environmental and Energy Policy," *Oxford Review of Economic Policy* Vol. 26, No. 2 (2010): 197–224. A selective survey of the literature to highlight what is known about the efficiency of particular kinds of policies, laws, and regulations in managing energy and environmental risk.

Griffin, James M., and Steven L. Puller. *Electricity Deregulation: Choices and Challenges* (Chicago, IL: University of Chicago Press, 2005). Bringing together leading experts from academia, government, and big business, this comprehensive volume provides a timely and thoughtful discussion of the many risks and rewards of electricity deregulation in practice.

Hill, J., E. Nelson, E. D. Tilman, S. Polasky, and D. Tiffany. "Environmental, Economic, and Energetic Costs and Benefits of Biodiesel and Ethanol Biofuels," *Proceedings of the National Academy of Sciences of the United States of America* Vol. 103, No. 30 (2006): 11206–11210. Available at http://www.pnas.org/cgi/doi/10.1073/pnas.0604600103.

International Energy Agency. *Renewable Energy: Market and Policy Trends in IEA Counties* (Paris: International Energy Agency, 2004). Examines policies and measures that have been introduced in IEA countries to increase the cost-effective deployment of renewables and evaluates the results.

Smith, J. L. "Inscrutable OPEC? Behavioral Tests of the Cartel Hypothesis," *The Energy Journal* Vol. 26, No. 1 (2005): 51–82.

Unger, T., and E. O. Ahgren. "Impacts of a Common Green Certificate Market on Electricity and CO_2-Emission Markets in the Nordic Countries," *Energy Policy* Vol. 33, No. 16 (2005): 2152–2163.

Additional References and Historically Significant References are available on this book's Companion Website: http://www.pearsonhighered.com/tietenberg/

8 Recyclable Resources: Minerals, Paper, Bottles, and E-Waste

Man is endowed with reason and creative powers to increase and multiply his inheritance; yet up to now he has created nothing, only destroyed. The forests grow ever fewer; the rivers parch; the wildlife is gone; the climate is ruined; and with every passing day the earth becomes uglier and poorer.

—Anton Chekhov, *Uncle Vanya*, Act I (1896)

Introduction

Once used, energy resources dissipate into heat energy. They cannot be recycled. Other resources, in contrast, retain their basic physical and chemical properties during use and under the proper conditions can be recycled or reused. They therefore represent a separate category for us to examine.

What is an efficient amount of recycling? Will the market automatically generate this amount in the absence of government intervention? How does the efficient allocation over time differ between recyclable and nonrecyclable resources? We begin our investigation by describing how an efficient market in recyclable, depletable resources would work. We then use this as a benchmark to examine recycling in some detail.

An Efficient Allocation of Recyclable Resources

Extraction and Disposal Cost

How would an efficient market, one devoid of any imperfections, allocate a recyclable depletable resource? The models developed in Chapter 6 provide a point of departure for answering this question. In the earliest periods, reliance would generally be exclusively on the virgin ore, because it is cheapest. As more concentrated ores are extracted, the mining industry would turn to the lower-grade ore and to foreign sources for higher-grade ores.

In the presence of technological progress, the increasing reliance on the lower-grade ores would not necessarily precipitate an increase in cost (as shown in Example 6.1), at least initially. Eventually, however, as the sources became increasingly difficult to extract, a point would be reached at which the costs of extraction and prices of the virgin material would begin to rise.

At the same time, the costs of disposing of the products would probably rise as population density became more pronounced and wealth levels supported higher levels of waste. Over the last two centuries, the world has experienced a large increase in the geographic concentration of people. The attraction of cities and exodus from rural areas led an increasingly large number of people to live in urban or near-urban environments.

This concentration creates waste disposal problems. Historically, when land was plentiful and the waste stream was less hazardous, the remnants could be buried in landfills. But as land became scarce, burial became increasingly expensive. In addition, concerns over environmental effects on water supplies and economic effects on the value of surrounding land have made buried waste less acceptable.

The rising costs of virgin materials and of waste disposal increase the attractiveness of recycling. By recovering and reintroducing materials into the system, recycling not only provides an alternative to virgin ores, but it also reduces the waste disposal load.

Consumers, as well as manufacturers, play a role on both the demand and supply sides of the market. On the demand side, consumers would find that products depending exclusively on virgin raw materials are subject to higher prices than those relying on the cheaper recycled materials. Consequently, consumers would have a tendency to switch to products made with the recycled raw materials, as long as quality is not adversely affected. This powerful incentive is called the *composition of demand effect*.

As long as consumers bear the cost of disposal, they have the additional incentive to return their used recyclable products to collection centers. By doing so they avoid disposal costs and reap financial rewards for supplying a product someone wants.

This highly stylized version of how the market should work has to be complemented by some hard realities that must be faced in setting up actual markets. For the cycle to be complete, it is essential that a demand exist for the recycled products. New markets may ultimately emerge, but the transition may prove somewhat turbulent. Simply returning recycled products to the collection centers accomplishes little if they are simply dumped into a nearby landfill or if the supply is increased so much by mandatory recycling laws that prices for recycled materials fall through the floor (thereby destroying the incentive to continue supplying them).

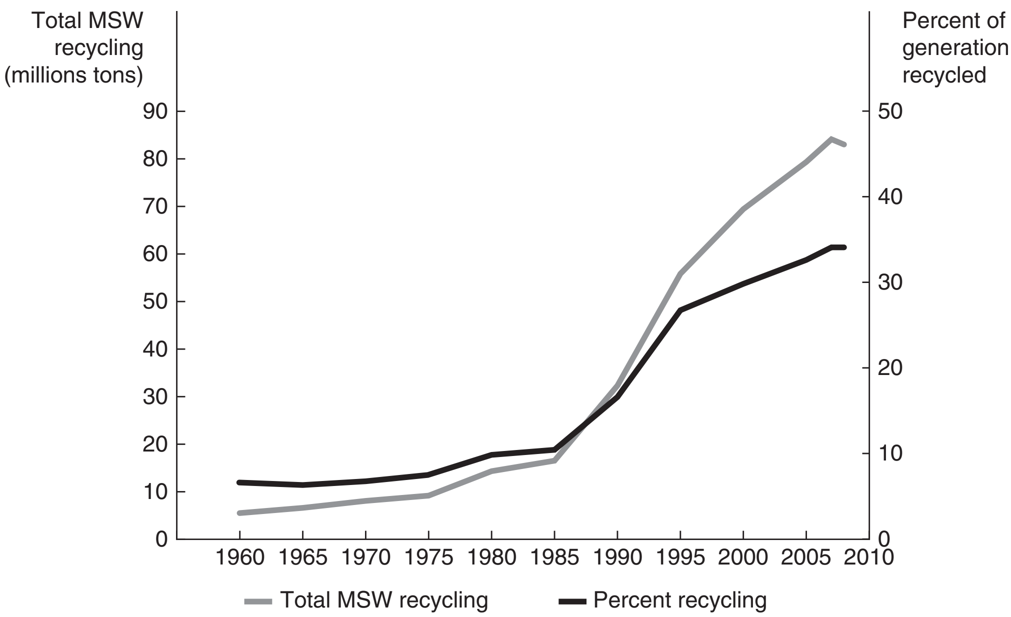
The purity of the recycled products also plays a key role in explaining the strength of demand for them. One of the reasons for the high rate of aluminum recycling and much lower rate of plastics recycling is the differential difficulty with which a high-quality product can be produced from scrap. Whereas bundles of aluminum cans have a relatively uniform quality, waste plastics tend to be highly contaminated with nonplastic substances or with plastics of very different types, and the plastics manufacturing process has little tolerance for impurities. Remaining contaminants in metals can frequently be eliminated by high-temperature combustion, but plastics are destroyed by high temperatures. Finally, waste that contains hazardous materials, such as mercury and lead, raises additional complexities. The rapidly growing stream of electronic waste (*e*-waste) contains both hazardous waste *and* valuable minerals, creating complicated dilemmas. As we discuss in a subsequent section of this chapter, markets for discarded electronics in industrializing countries may lack good enforcement mechanisms to ensure proper disposal of the hazardous components.

Recycling: A Closer Look

The model in the preceding section would lead us to expect that recycling would increase over time as virgin ore and disposal costs rose. This seems to be the case. Take copper, for example. During 1910, recycled copper accounted for about 18 percent of the total production of refined copper in the United States. By 2004 this percentage had risen to 29 percent. Approximately 40 percent of the world's copper demand today is met by recycling. And, according to the Bureau of International Recycling, an estimated 70 percent of the copper scrap exported by the United States is used by industries in China.

For other materials, recycling rates are on the rise. According to the U.S. EPA, the rate of recycled waste in the United States has risen to 33.2 percent in 2008 (Figure 8.1). For certain materials, the rates are even higher (99 percent for auto batteries, 71 percent for office paper, 48 percent for aluminum beer and soft

FIGURE 8.1 Municipal Solid-Waste Recycling Rates 1960–2008



Source: "Municipal Solid Waste Generation, Recycling, and Disposal in the United States: Facts and Figures for 2008," Figure 2. United States Environmental Protection Agency.

drink cans, 63 percent for steel packaging). Plastic polyethylene terephthalate (PET) bottles were recovered at a rate of 37 percent, high-density polyethylene (HDPE) bottles at a rate of 28 percent, and glass containers at a rate of 28 percent. According to another indicator, by 2006, over 8,600 curbside recycling programs were in existence. Twenty years earlier, only one curbside program was in place.

In most cases recycling is not cheap. While several types of costs are involved, transport and processing costs are usually especially significant. The sources of scrap may be concentrated around cities where most of the products are used, while for historical reasons the processing facilities are near the sources of the virgin ores. The scrap must be transported to the processing facility and the processed scrap to the market.

Labor costs are an important component of the processing costs. Collecting, sorting, and processing scrap is typically very labor intensive. Higher labor costs can make the recycled scrap less competitive in the input market. Recognizing the importance of labor costs raises the possibility that recycling rates would be higher in regions where labor costs are lower, which does seem to be the case. For example, Porter (1997) shows how vibrant markets for scrap have emerged in Africa.

Energy costs also matter. According to the Bureau of International Recycling (BIR), recycling offers significant energy savings over production from raw materials. For example, steel recycling expends 74 percent, aluminum 95 percent, copper 85 percent, paper 64 percent, and plastics 80 percent less energy. Additionally, producing materials via recycling results in less water and air pollution. BIR estimates that the production of paper via recycling causes 35 percent less water pollution and 74 percent less air pollution.

And, finally, since the processing of scrap as input into the production process can produce its own environmental consequences, compliance with environmental regulations can add to the cost of recycled input. In the United States, for example, relatively low world copper prices, coupled with high environmental compliance costs, created a cost squeeze that contributed to the closure of all U.S. secondary smelters and associated electrolytic refineries by 2001.

When recycling markets operate smoothly however, scrap becomes a cost-competitive input, and rather dramatic changes occur in the manufacturing process. Not only do manufacturers rely more heavily on recycled inputs, but also they begin to design their products to facilitate recycling. Facilitating recycling through product design is already important in industries where the connection between the manufacturer and disposal agent is particularly close. Aircraft manufacturers, which are often asked to scrap old aircraft, may stamp the alloy composition on parts during manufacturing to facilitate recycling. The idea is beginning to spread to other industries. Ski boot manufacturers in Switzerland, for example, are beginning to stamp all individual boot parts with a code to identify their composition.

Recycling and Ore Depletion

How does the efficient allocation of a recyclable resource compare with that of a nonrecyclable resource over time? Thinking back to the models in Chapter 6, perhaps the most important difference occurs in the timing of the switch point.

As long as the resource can be recycled at a marginal cost lower than that of the substitute, the market tends to rely on the recyclable resource longer than it does on a nonrecyclable resource with an identical extraction cost curve. This should not be surprising, since one effect of recycling is simply to add more of the resource.

This point can be illustrated using a simple numerical example. Suppose 100 units of a resource are contained in a product with a useful life of one year. Suppose further that 90 percent of the resource could be recovered and reused after one year. During the first year, the full 100 units could be used. At the end of the second year, 90 percent of the remaining 90 units could once again be recovered, leaving 81 units for the third year, and so on.

How much more of this resource was made available by recycling? Algebraically, if we let the original stock be A, and the recovery rate be a, then the total amount used would be an infinite sum of the form $A + Aa + Aa^2 + Aa^3$. It turns out that the sum of this series as time becomes infinitely long is $A/(1 - a)$.[1] Notice that nonrecyclable resources are represented by the special case where $a = 0$. In this case the sum of resource use equals the available stock. Whenever $a > 0$, however, as it would be when any of the resource was recycled, the sum of the resource flows exceeds the size of the original stock. The closer to 1.0 a is, the larger the sum of the resource flows. For example, if $a = 0.9$, as it was in our example, the sum of the flows is 10 times the size of the stock. The effect of recycling is to increase the size of the available resources by a factor of 10.

This formulation also points out another feature of recycling. Unless the recycling rate is 100 percent ($a = 1.0$), the sum of the resource flows is finite. This means that while some recycled materials can be recycled forever, the amount will become infinitesimally small as time goes on.

An efficient economic system will orchestrate a balance between the consumption of newly mined and recycled materials, between disposing of used products and recycling, and between imports and domestic production. Example 8.1 provides an example of how changing economic circumstances can lead to an increase in recycling.

Factors Mitigating Resource Scarcity

Recycling is promoted by resource scarcity, but resource scarcity is, in turn, affected by a number of other factors. Three alternatives have been particularly important: (1) exploration and discovery, (2) technological progress, and (3) substitution.

Exploration and Discovery

A profit-maximizing firm will undertake exploration activity until the marginal discovery cost equals the marginal scarcity rent received from a unit of the

[1]Note the similarity of $1/(1 - a)$ to the familiar multiplier used in introductory macroeconomics, $1/(1 - MPC)$.

EXAMPLE 8.1

Lead Recycling

The domestic demand for lead has changed significantly over the last 30 years. In 1972 dissipative, nonrecyclable uses of lead (primarily gasoline additives, pigments in paint, and ammunition) accounted for about 30 percent of reported consumption. And only about 30 percent of all produced lead came from recycled material.

Over the last three decades, however, congressional recognition of lead's negative health effects on children has led to a series of laws limiting the amount of allowable lead in gasoline and paints. This has resulted not only in a decline in the total amount of lead used, but also in the dramatic decline of the dissipative uses (which, by 1997, had fallen to only 13 percent of total demand). A declining role for dissipative uses implies that an increasing proportion of the production is available to be recycled. And, in fact, more is now recycled. By 2010, 80 percent of the domestic lead consumption came from recycled scrap. The lead-acid battery industry continues to be the largest user of lead.

Old (postconsumer) scrap accounts for nearly all the total lead scrap recovered. Used batteries supply about 90 percent of that old scrap. Battery manufacturers have begun entering buyback arrangements with retail outlets, both as a marketing tool for new batteries and as a means of ensuring a supply of inputs to their downstream manufacturing operations. Contrast this with aluminum, for example. In 2006, 64 percent of recycled aluminum came from new (manufacturing) scrap, while only 36 percent was from old scrap (beverage cans and other discaded aluminum products).

Source: U.S. Department of the Interior. *Minerals Yearbook* available on the Web at: http://minerals.usgs.gov/minerals/pubs/mcs/

resource sold.[2] Since the marginal scarcity rent—the difference between the price received and the marginal cost of extraction—is the marginal benefit received by the firm engaging in exploration activity, the level of activity should be increased to maximize profits until this marginal benefit is equal to the marginal cost.

An understanding of this relationship between scarcity rent and marginal discovery cost allows us to think about how exploration activity would respond to population and income growth. Since both of these factors contribute to rising demand over time, they raise the marginal user cost and the scarcity rent, stimulating producers to undertake larger marginal discovery costs.

How much this demand pressure is relieved depends upon the amount of exploration activity and the amount of resources discovered per unit of exploration activity undertaken. If the marginal discovery cost curve is flat (implying a large amount of relatively available resources), increases in scarcity rent can stimulate

---

[2]This is strictly true only when no uncertainty is associated with exploration. Even with uncertainty, however, marginal discovery cost is highly related to scarcity rent. See Devarajan and Fisher (1982).

large amounts of successful exploration activity. If the marginal discovery cost curve is steeply sloped (as would be the case when exploration had to take place in increasingly hostile and unproductive environments), increases in scarcity rent stimulate less exploration activity.

Technological Progress

Technological progress reduces the cost of ore by discovering new ways to extract, process, and use it. In Chapter 6, for example, we showed the significant impact of pelletization on the cost of producing steel from iron ore. The effect was so dramatic that production costs actually fell over time in spite of the need to use a lower-grade ore.

It is important to realize that the rate and type of technological progress are influenced by the degree of resource scarcity. Rising extraction costs create new profit opportunities for the development of new technologies. These profit opportunities are largest for technologies that economize on scarce resources and utilize abundant ones. In periods when labor is scarce and capital abundant, new technologies tend to use capital and save labor. If population growth were to reverse the relative scarcity, subsequent technological progress would concentrate on using labor and saving capital. In the past, when fossil-fuel energy was abundant and cheap, newly discovered technologies relied heavily on this energy source. As fossil-fuel supplies decline, technological progress can be expected to economize by increasing the amount of useful energy received per unit of fossil-fuel input and by replacing fossil-fuel energy with forms of renewable energy.

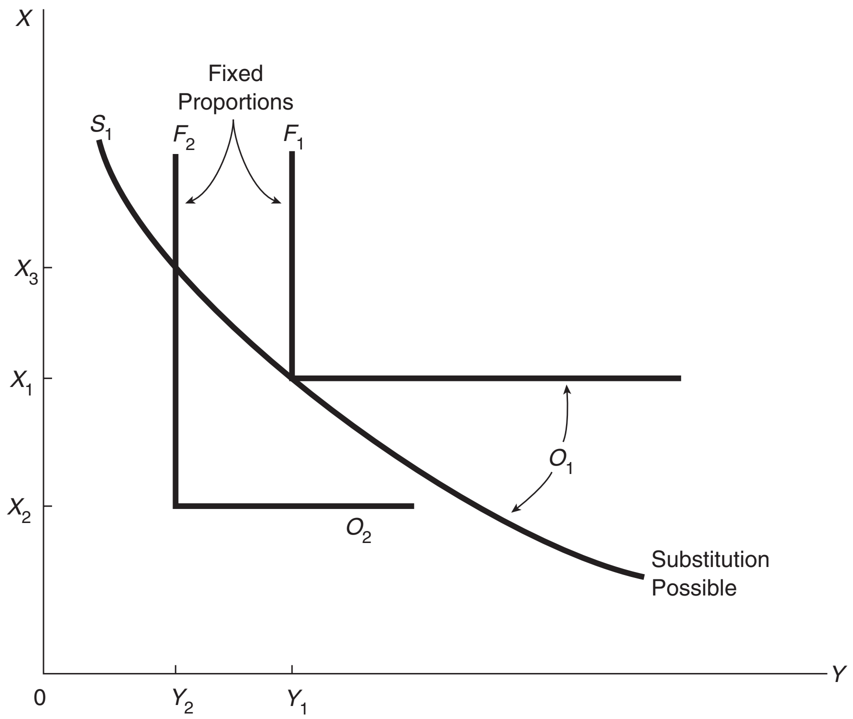
Substitution

The final way in which adverse consequences of resource scarcity can be mitigated is by substituting abundant resources for scarce ones. The easier the substitution of abundant depletable or renewable resources, the smaller will be the impact of declining availability and rising costs (see Figure 8.2).

In the graph, three isoquants (S_1, F_1, F_2) are plotted. An *isoquant* portrays all the possible combinations of inputs that can produce a given level of output. The two right-angled isoquants (F_1 and F_2) depict the fixed-proportions case, the case in which no input substitution is possible. The fixed-proportions isoquant nearer the origin (F_2) refers to a lower output level than the other fixed-proportion isoquant (F_1). The third isoquant (S_1) does show some possibility for input substitution and is drawn in such a way as to produce the same output level (O_1) as F_1. Naturally it implies a different production technology or set of technologies from F_1.

We can illustrate the significance of input substitution on output using Figure 8.2. Assume that the amount of some input Y (a depletable resource) is reduced from Y_1 to Y_2. If the technology involved is characterized by S_1, the constant output level (O_1) can be maintained by increasing the amount of the other resource

FIGURE 8.2 Output Levels and the Possibilities for Input Substitution



Source: Adapted from Table 3.9 in J. B. Opschoor and Dr. Hans B. Vos, *Economic Instruments for Environmental Protection* (Paris: Organization for Economic Cooperation and Development), p. 53.

used from X_1 to X_3. This increase in X compensates for the reduction in Y, leaving output unaffected.

Notice what happens, however, when the production process is characterized by F_1 instead of S_1. A reduction in the availability of Y from Y_1 to Y_2 necessitates a reduction in output from O_1 to O_2. No substitution of X for Y is possible. In addition, because inputs must be used in fixed proportions, the amount of X would be reduced from X_1 to X_2. Any more X would be redundant; it would not result in any additional output. These examples serve to illustrate a basic premise—the wider the array of substitution possibilities, the smaller the impact of resource scarcity on output.

This short review suggests that some factors (e.g., rising population and incomes) increase the likelihood of resource scarcity, while others (e.g., exploration and discovery, technological progress, and input substitution) mitigate the seriousness of scarcity. If resource scarcity is increasing in some sense, we should be able to discover that natural resource prices are rising more rapidly than prices in general (see Example 8.2). Resource prices, in turn, affect incentives to recycle as do the marginal costs of disposal.

EXAMPLE 8.2

The Bet

In 1980 each of two distinguished protagonists in the scarcity debate "put his money where his mouth is." Paul Ehrlich, an ecologist with a strong belief in impending scarcity, answered a challenge from Julian Simon, an economist known for his equally strong belief that concerns about impending scarcity were groundless. According to the terms of the bet, Ehrlich would hypothetically invest $200 in each of any five commodities he selected. (He picked copper, chrome, nickel, tin, and tungsten.) Ten years later the aggregate value of the same amounts of those five commodities would be calculated in real terms (after accounting for normal inflation). If the value increased, Simon would send Ehrlich a check for the difference. If the value decreased, Ehrlich would send Simon a check for the difference.

In 1990 Ehrlich performed the calculations and sent Simon a check for $576.07. Real prices for each of the five commodities were lower; some were less than half their former levels. New sources of the minerals had been discovered, substitutions away from these minerals had occurred in many of their uses (particularly computers), and the tin cartel, which had been holding up tin prices, collapsed.

Would the outcome of the Simon-Ehrlich wager have been the same if the bet had covered the entire twentieth century? According to a subsequent analysis of the data on these same minerals by McClintock and Emmett (2005), despite ups and downs in prices over the course of the past century, Simon would also have won even a century-long wager.

Finally, how would Simon have fared in decades other than the one covered by the bet? Was he just lucky to have picked the 1980s? It turns out that to some extent he was lucky. Of the ten decades in that century he would have won in five decades (the 1900s, 1910s, 1940s, 1980s, and 1990s) and lost in the remaining five. He would have lost by a few dollars in the 1950s and by more significant amounts in the other four decades. Does this evidence provide a lesson for the future? You be the judge.

Sources: John Tierney, "Betting the Planet." THE NEW YORK TIMES MAGAZINE, December 2, 1990, pp. 52–53, 74, 76, 78, 80–81; and D. McClintick and Ross B. Emmett, "The Simon-Ehrlich Debate." PERC Reports 23 (2005) (3), pp. 16–17.

Market Imperfections

As we discovered in the discussion of the role of oil in national security, when mineral imports are critically important and come from risky sources, the market perceives a biased price ratio, one that fails to incorporate some of the social costs of imports. The result would be an inefficient and excessive reliance on imports.

Other market imperfections are apparent as well. An unbalanced treatment of waste by producers and consumers can lead to biases in the market choices between recycling and the use of virgin ores. Since disposal cost is a key ingredient in determining the efficient amount of recycling, the failure of an economic agent to bear the full cost of disposal implies a bias toward virgin materials and away from recycling. We begin by considering how the method of financing the disposal of potentially recyclable waste affects the level of recycling.

Disposal Cost and Efficiency

The efficient level of recycling depends on the marginal cost of disposal. Suppose, for example, it costs a community $20 per ton to recycle a particular waste product that can ultimately be sold to a local manufacturer for $10 per ton. Can we conclude that this is an inefficient recycling venture because it is losing money? No, we can't! In addition to earning the $10 per ton from selling the recycled product, the town is avoiding the cost of disposing of the product. This avoided marginal cost is appropriately considered a marginal benefit from recycling. Suppose the marginal avoided disposal cost was $20 per ton. In this case, the benefits to the town from recycling would be $30 per ton ($20 per ton avoided cost plus $10 per ton resale value) and the cost would be $20 per ton; this would be an efficient recycling venture. Both marginal disposal costs and the prices of recycled materials directly affect the efficient level of recycling.

The Disposal Decision

Potentially recyclable waste can be divided into two types of scrap: old scrap and new scrap. *New scrap* is composed of the residual materials generated during production. For example, as steel beams are formed, the small remnants of steel left over are new scrap. *Old scrap* is recovered from products used by consumers.

To illustrate the relative importance of new scrap and old scrap, consider that in the U.S. aluminum industry, about 40 percent of the recovered aluminum scrap comes from old scrap. The difficulties in recycling new scrap are significantly less than those in recycling old scrap. New scrap is already at the place of production, and with most processes it can simply be reentered into the input stream without transportation costs. Transport costs tend to be an important part of the cost of using old scrap.

Equally important are the incentives involved. Since new scrap never leaves the factory, it remains under the complete control of the manufacturer. Having the joint responsibility of creating a product and dealing with the scrap, the manufacturer now has an incentive to design the product with the use of the new scrap in mind. It would be advantageous to establish procedures guaranteeing the homogeneity of the scrap and minimizing the amount of processing necessary to recycle it. For all these reasons, it is likely the market for new scrap will work efficiently and effectively.

Unfortunately, the same is not true for old scrap. The market works inefficiently because the product users do not bear the full marginal social costs of disposing of their product. As a result, the market is biased away from recycling old scrap and toward the use of virgin materials.

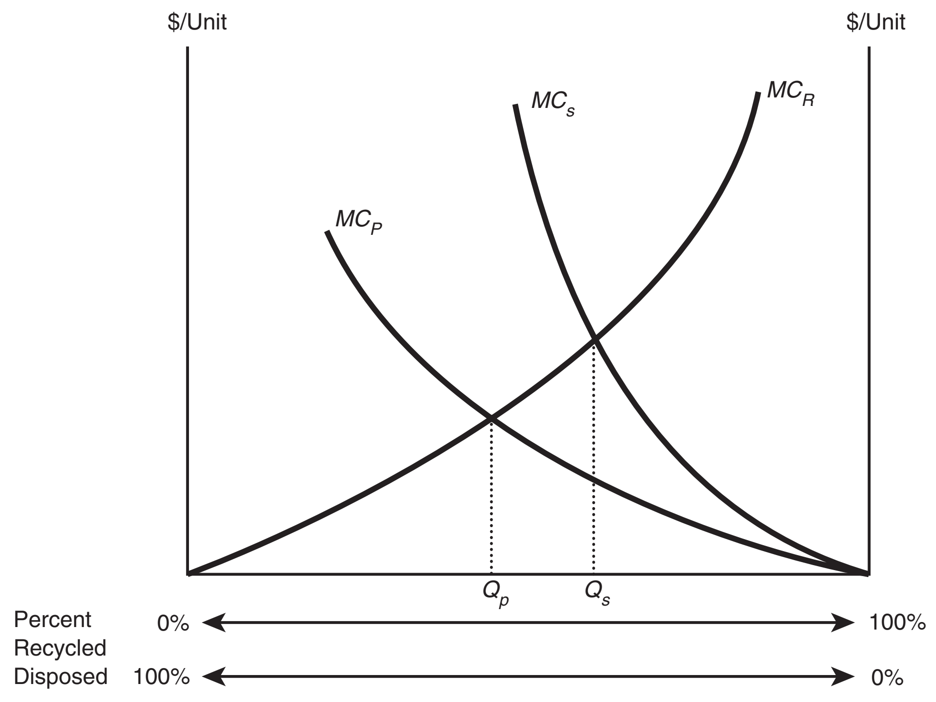
The key to understanding why these costs are not internalized lies in the incentives facing individual product users. Suppose you had some small aluminum products that were no longer useful to you. You could either recycle them, which usually means driving to a recycling center, or you could toss them into your trash. In comparing these two alternatives, notice that recycling imposes one cost on you (transport cost), while the second imposes another (disposal cost).

It is difficult for consumers to act efficiently because of the way trash collection has traditionally been financed. Urban areas have generally financed trash collection with taxes, if publicly provided, or a flat-rate fee, if privately provided. Neither of these approaches directly relates the size of an individual's payment to the amount of waste. The *marginal* cost to the homeowner of throwing out one more unit of trash is negligible, even when the cost to society is not. The marginal private disposal cost and the cost to society as a whole diverge (see Figure 8.3).

When the private marginal cost of disposal (MC_P) is lower than the marginal social cost of disposal (MC_s), the market level of recycling [where the marginal cost of recycling (MC_R) is equal to the marginal private disposal cost] is inefficient. Only if all social costs are included in the marginal cost of disposal will the efficient amount of recycling (Q_s) be attained.[3]

This point can be reinforced by a numerical example. Suppose your city provides trash pickup for which you pay $150 a year in taxes. Your cost will be

FIGURE 8.3 The Efficient Level of Recycling



[3]According to Figure 8.3, would 100 percent recycling normally be efficient? Does that conclusion make sense to you? Why or why not?

$150 regardless (within reasonable limits) of how much you throw out. In that year your additional (marginal) cost from throwing out these items is *zero*. Certainly the marginal cost to society is not zero, and, therefore, the balance between these alternatives as seen by the individual homeowner is biased in favor of throwing things out.[4]

Littering is an extreme example of what we have been talking about. In the absence of some kind of government intervention, the cost to society of littering is the aesthetic loss plus the risk of damage to automobile tires and pedestrians caused by sharp edges of discarded cans or glass. Tossing used containers outside the car is relatively costless for the individual, but costly for society.[5]

Disposal Costs and the Scrap Market

How would the market respond to a policy forcing product users to bear the true marginal disposal cost? The major effect would be on the supply of materials to be recycled. Recycling would now offer consumers a way to avoid disposal costs and possibly even be paid for discarded products. This would cause the diversion of some materials to recycling centers, where they could be reintegrated into the materials process. If this expanded supply allows dealers to take advantage of previously unexploited economies of scale, this expansion could well result in a lower average cost of processing, as well as more recycled materials.

The total consumption of inputs would increase because the price falls. The use of recycled materials increases as well. The amount of virgin ore falls. Thus, the correct inclusion of disposal cost would tend to increase the amount of recycling and extend the useful economic life for depletable, recyclable resources.

Subsidies on Raw Materials

Disposal costs are only part of the story. Inputs derived from recycling can only compete with raw materials if the playing field is level. Subsidies on raw materials are another troubling source of inefficiencies that create a bias away from recycled inputs.

Subsidies can take many forms. One form is illustrated by the U.S. Mining Law of 1872. This law, which was originally passed more than 150 years ago to promote mining on public lands, is still on the books. Under this law, miners can stake lode claims (for subsurface minerals) and placer claims (for surface minerals) for mineral prospecting on public lands. A claim can be maintained for a payment of only $100 a year. If minerals are discovered in a claim area and at least $500 has been invested in development or extraction, the land could actually be bought for $5 an acre on lode claims or $2.50 an acre on placer claims. In 1999 the U.S. Congress enacted a moratorium on land purchases, but not on staking claims.

---

[4]The problem is not that $150 is too low; indeed it may be too high! The point is that the cost of waste disposal does not increase with the amount of waste to be disposed.

[5]Using economic analysis, would you expect transients or residents to have a higher propensity to litter? (Why?)

These prices for access to public lands are so low relative to market prices that they constitute a considerable implicit subsidy. As a result of this subsidy, taxpayers not only don't receive the true value of the mining services provided by public lands, but the subsidy has the effect of lowering the cost of extracting these raw minerals. As a result, raw materials are artificially cheap and can inefficiently undermine the market for recycled inputs.

Corrective Public Policies

Why are recycling rates so low? No doubt some of the responsibility lies in improper incentives created by inappropriate pricing. Can the misallocation resulting from inefficiently low disposal cost be corrected?

One approach, volume pricing, would impose disposal charges reflecting the true social cost of disposal (see Example 8.3).

A preimplementation concern about volume pricing was that it might impose a hardship on the poor residents of the area. Strategies based on higher prices always raise the specter that they will end up placing an intolerable burden on the poor. In the case examined by Example 8.3 that concern was apparently misplaced. Under the old system of financing trash collection, every household paid the same fee regardless of how much trash is produced. Since lower-income

EXAMPLE 8.3

Pricing Trash in Marietta, Georgia

In 1994 the people of Marietta, Georgia participated in a demonstration project that changed the way in which waste was priced. The traditional $15 monthly fee for trash pickup was cut to $8 per month. In addition, half of the residents faced a per-bag price on waste ($0.75 per bag), while the rest faced a monthly fee for pickup that depended on the maximum number of cans per month that the customer wished to have picked up per month. This number was contracted in advance by the customer and did not vary from month to month. The fee was $3 or $4 per can (depending upon the number).

Economic theory suggests that while both plans should reduce waste and increase recycling, the per-bag fee should promote more. (Can you see why?)

And indeed that is what happened. The can program reduced nonrecycled waste by about 20 percent, whereas the bag program reduced it by as much as 51 percent. Both programs had an equally strong effect on encouraging households to recycle. Both programs not only diverted waste into recycling, they also reduced the amount of waste generated.

Could the costs associated with the program be justified in benefit–cost terms? According to the economists who conducted the study, they were. The net benefits for the city were estimated to be $586 per day for the bag program and $234 per day for the can program.

Source: G. L. Van Houtven and G. E. Morris. "Household Behavior Under Alternative Pay-as-You-Throw Systems for Solid Waste Disposal," *Land Economics* Vol. 75, No. 4 November 1999, pp. 515–537.

households produce less trash, they were, in effect, subsidizing wealthier households. Under the new system, lower-income households pay only a flat fee since they don't need to purchase stickers for additional disposal. The expense of these stickers is less than the average cost of disposal, which was the basis for the previous fee. Poor households have turned out to be better off, not worse off, under the new pricing system.

Another suggestion for promoting recycling now being applied in many areas is the refundable deposit. Already widely accepted for beverage containers, such deposits could become a remedy for many other products.

A refund system is designed to accomplish two purposes: (1) the initial charge reflects the cost of disposal and produces the desired composition of demand effect; and (2) the refund, attainable upon turning the product in for recycling, helps conserve virgin materials. Such a system is already employed in Sweden and Norway to counter the problem of abandoned automobiles.

The recycling of aluminum beverage cans has been one clear beneficiary of deposit refund schemes.[6] Quite a few countries, including Germany, Finland, Norway, Denmark, Sweden, Barbados, Canada, and the state of South Australia, and 11 U.S. states have container deposit refund programs in place. Although not all states have passed bottle bills, over 50 percent of aluminum beverage cans are now recycled in the United States. As a result, aluminum old scrap has become an increasingly significant component of total aluminum supplies. Recycling aluminum saves about 95 percent of the energy that is needed to make new aluminum from ore. The magnitude of these energy savings has had a significant influence on the demand for recycled aluminum as cost-conscious producers search for new ways to reduce energy costs. Debate 8.1 explores why only some states have implemented refundable bottle deposits.

Beverage container deposits also reduce illegal disposal (littering) because an incentive is created to bring the bottle or can to a recycling center. In some cities, scavenging and returning these bottles has provided a significant source of income to the homeless. One Canadian study found that recycling creates six times as many jobs as landfilling.

Deposit-refund systems are also being used for batteries and tires. New Hampshire and Maine, for example, place a surcharge on new car batteries. Consumers in these states receive a rebate if they trade in their used battery for a new one. Oklahoma places a $1 fee on each new tire sold and then returns $0.50 to certified processing facilities for each tire handled.

Some states in the United States, as well as some developing countries, also use deposit-refund systems to assure that pesticide containers are returned after use. Since these containers usually contain toxic residues after use, which can contaminate water and soil, collecting the containers and either reusing them or properly decontaminating them can eliminate this contamination threat.

Some areas attempt to enlist economic incentives by imposing a disposal or recycling surcharge on the product. Paid at the time of purchase of a new product,

---

[6]A very strong demand for aluminum scrap was also influential. In fact, the price for aluminum scrap went so high in 1988 that pilferers were stealing highway signs and guardrails for their aluminum content.

DEBATE 8.1

"Bottle Bills": Economic Incentives at Work?

Ten U.S. states—California, Connecticut, Hawaii, Iowa, Maine, Massachusetts, Michigan, New York, Oregon, and Vermont—have passed "bottle bill" legislation. One city, Columbia, Missouri, also passed legislation, but it was repealed in 2002. Delaware's bottle-deposit system was repealed in 2010, effective February 2011. Every year, several states either have proposed new legislation or proposed expansions of existing legislation. More often than not, these proposed bills do not pass. Bottle deposits in the United States range from $0.05 to $0.15 and laws vary on which containers are redeemable for deposits.

While on average, U.S. container recycling rates have been below 40 percent, recycling rates in bottle-deposit states are much higher, averaging around 80 percent. Michigan's $0.10 beverage can deposit produced recycling rates close to 100 percent. Statistics on litter reduction show the largest gains in bottle-deposit states.

Although bottle-deposit states have recycling rates double those of states without deposits, that is not sufficient evidence to suggest that it would be efficient for all states to have them.

Economic studies on the efficiency of bottle deposits are limited. Porter (1983) estimated the costs and benefits of the then newly passed Michigan bottle bill. He found that for most estimates of costs and benefits, the bill passed a benefit–cost test. Ashenmiller (2009) finds that bottle deposits increase the numbers of recycled containers and reduce waste stream costs by diverting these containers away from curbside programs. Using survey data from California, he finds between 36 percent and 51 percent of materials at redemption centers would not have been collected using existing curbside programs alone (without the complementary deposit-refund system). Interestingly, however, some of the success of the California program can be attributed to its design—its curbside programs use volume-based pricing for trash. This analysis also notes that curbside programs work best in densely populated areas and that cash recycling programs can be an important income source for the working poor.

Since the efficiency of deposit–refund systems depends on their cost, they may be efficient for some states, but not others. Key determinants of the relative costs of bottle deposits vary from state to state. Disposal costs depend on landfill availability, and return rates depend on population densities and distances to redemption centers. States with bottle deposits may incur the extra expense of illegal returns from bottles purchased in nearby states that do not require a deposit. Enforcement across state lines is costly and imperfect. States with large bottlers like Coca-Cola are usually opposed to bottle deposits. Does your state have a bottle deposit? Does that seem the right choice?

Sources: http://globalwarming.house.gov/mediacenter/pressreleases?id=0126; www.containerrecyclinginstitute.org; Richard C. Porter, "Michigan's Experience with Mandatory Deposits on Beverage Containers." LAND AND ECONOMICS, 59 (1983); Bevin Ashenmiller, "Cash Recyling, Waste Disposal Costs, and the Incomes of the Working Poor: Evidence from California." LAND ECONOMICS, 85(3), August 2009.

this surcharge would normally be designed to recover the costs of recycling the product at the end of its useful life; more-difficult-to-recycle products would have larger fees. These fees would normally be coupled with a requirement that the revenue be used by sellers to set up recycling systems. Assuming these fees correctly internalize the costs of recycling, they will provide consumers with incentives to take the recycling and disposal costs into account, since easier-to-recycle products would have a lower price (including the fee). Note, however, that these recycling surcharges do not provide any incentive against illegal disposal, since the consumer gets no rebate for dropping the product off at a collection center, but it they provide no specific incentive for illegal disposal either. Since the surcharge is paid up front, it cannot be avoided by illegal disposal. In this sense, the deposit-refund system is clearly superior to either recycling surcharges or volume pricing of trash.

The tax system can also be used to promote recycling by taxing virgin materials and by subsidizing recycling activities. The European approach to waste oil recycling, reinforced by the high cost of imported crude oil, was to require both residential and commercial users to recycle all waste oil they generate. Virgin lubricating oils are taxed, and the resulting income is used to subsidize the recycling industry. As a result, many countries collect up to 65 percent of the available waste oil.

In the United States, which does not subsidize waste oil recycling, the waste oil market has been rather less successful, but it is growing. In California in 2005, almost 60 percent of used lubricated oil was recycled. Laws in most states prohibit used oil disposal.

Many areas are now using tax policy to subsidize the acquisition of recycling equipment in both the public and private sectors. Frequently taking the form of sales-tax exemptions or investment tax credits to private industries or loans or grants to local communities, these approaches are designed to get recycling programs off the ground, with the expectation that they will ultimately be self-sustaining. The pioneers are being subsidized.

Examining Oregon's program can serve to illustrate how a tax approach works. From 1981 to 1987, to reduce energy consumption as well as to promote recycling, the Oregon Department of Energy granted tax credits to 163 projects. Being granted this credit allows companies a five-year period in which to deduct from their taxes an amount equal to 35 percent of the cost of any equipment used solely for recycling. Oregon also offers a broader tax credit that covers equipment, land, and building purchases. Paper companies, the major recipients of both types of credits, have used them to increase the capacity to use recycled newsprint and cardboard in the papermaking process. According to Shea (1988), these incentives helped to raise Oregon's newspaper recycling rate (65 percent) to twice the national average.

Any long-run solution to the solid-waste problem must not only influence consumer choices about purchasing, packaging, and disposal, it must also influence producer choices about product design (to increase recyclability), product packaging, and the use of recycled (as opposed to virgin materials) in the production process. One general approach is called extended producer responsibility, and it involves requiring producers to take back packaging, and even their products, at the end of their useful life (see Example 8.4).

EXAMPLE 8.4

Implementing the "Take-Back" Principle

According to the "take-back" principle, all producers should be required to accept responsibility for their products—including packaging—from cradle to grave by taking them back once they have outlived their useful lives. In principle, this requirement was designed to encourage the elimination of inessential packaging, to stimulate the search for products and packaging that are easier to recycle, and to support the substitution of recycled inputs for virgin inputs in the production process.

Germany has required producers (and retailers as intermediaries) to accept all packaging associated with products, including such different types of packaging as the cardboard boxes used for shipping hundreds of toothbrushes to retailers, to the tube that toothpaste is sold in. Consumers are encouraged to return the packaging by means of a combination of convenient drop-off centers, refundable deposits on some packages, and high disposal costs for packaging that is thrown away.

Producers responded by setting up a new, private, nonprofit corporation, the Duales System Deutschland (DSD), to collect the packaging and to recycle the collected materials. This corporation is funded by fees levied on producers. The fees are based on the number of kilograms of packaging the producers use. The DSD accepts only packaging that it has certified as recyclable. Once certification is received, producers are allowed to display a green dot on their product, signaling consumers that this product is accepted by the DSD system. Other packaging must be returned directly to the producer or to the retailer, who returns it to the producer.

The law has apparently reduced the amount of packaging produced and has diverted a significant amount of packaging away from incineration and landfills. A most noteworthy failure, however, was the inability of the DSD system to find markets for the recycled materials it collected. Some German packaging even ended up in neighboring countries, causing some international backlash. The circumstance where the supply of recycled materials far exceeds the demand is so common—not only in Germany, but in the rest of the world as well—that further efforts to increase the degree of recycling will likely flounder unless new markets for recycled materials are forthcoming.

Despite the initial difficulties with implementing the "take-back" principle, the idea that manufacturers should have ultimate responsibility for their products has a sufficiently powerful appeal that it has moved beyond an exclusive focus on packaging and is now expanding to include the products themselves. In 2002, the European Union (EU) passed a law that makes manufacturers financially responsible for recycling the appliances they produce. In 2004, the European Union's Waste Electrical and Electronic Equipment (WEEE) directive came into effect, making it the responsibility of the manufacturers and importers in EU states to take back their products and to properly dispose of them.

As part of the WEEE program, a pilot study was conducted in Beijing, Delhi, and Johannesburg. This study found that e-waste recycling has developed in all three countries as a market-based activity.

Sources: A. S. Rousso and S. P. Shah. "Packaging Taxes and Recycling Incentives: The German Green Dot Program," *National Tax Journal* Vol. 47, No. 3 (September 1994): 689–701; Meagan Ryan. "Packaging a Revolution," *World Watch* (September–October 1993): 28–34; Christopher Boerner and Kenneth Chilton. "False Economy: The Folly of Demand-Side Recycling," *Environment* Vol. 36, No. 1 (January/February 1994): 6–15; R. Widmer, H. Oswald-Krapf, D. Sinha-Khetriwal, M. Schnellmann, and H. Boni. "Global Perspectives on E-waste," *Environmental Impact Assessment* Vol. 25 (2005): 436–458.

Markets for Recycled Materials

Successful recycling programs depend to a large extent on the existence of markets (buyers) for recycled materials. Consider plastics, for example. Currently, PET bottles are primarily used in carpet fiber and textiles including fleece. Other potential future uses for recycled PET bottles include waterproof shipping containers and coating for paper. HDPE plastics are primarily made into bottles and garden products, such as lawn edging and lawn chairs.

The market for plastics is expanding in some areas where the capacity to process the postconsumer waste and the demand for that material is greater than the amount recovered. As new uses expand, this market can be expected to grow. According to the U.S. EPA, the American Society for Testing and Materials (ASTM) is using new test methods that are facilitating the use of recycled plastics in building materials.

E-Waste

Recognizing the dangers from improperly disposed electronic equipment (e-waste), some states have enlisted economic incentives to promote recycling. The EPA reports that in 2006–2007, 2.5 million tons of TVs, computers, computer accessories, and cell phones were discarded; 82 percent (1.84 million tons) were discarded in landfills. This does not include the approximately 235 million units (not including cell phones) that have accumulated in storage (as of 2007). Although this represented less than 2 percent of the municipal solid-waste stream, electronics waste is a fast-growing segment of it, bringing with it rising concerns about the environmental and health effects of some of this waste. Lead, mercury, cadmium, and brominated flame retardants are all widely used in electronics. All of these substances have been linked to health risks, especially for children, and are considered hazardous waste.

Did you upgrade your cell phone this year? Perhaps you got a new laptop for school. Or maybe you got a new MP3 player or iPad. What happened to the old one? The US Geological Survey (USGS) reports some 1 billion cell phones were in use worldwide in 2002. In the United States alone, the USGS reports that cell phone subscribers increased from 340,000 in 1985 to 180 million in 2004. Moreover, 130 million cell phones were retired in 2005. The U.S. EPA reports that currently less than 1 percent of cell phones discarded are recycled. Are the minerals used to make cell phones (primarily copper, iron, nickel, silver, and zinc, with smaller amounts of aluminum, gold, lead, palladium, and tin) valuable? Apparently quite valuable! Table 8.1 shows the estimated value of the metals in cell phones in the United States.

Table 8.2 lists the 24 states with e-waste legislation. To take an early example of this legislation, California passed a bill in 2003 that charges consumers a fee for buying computer monitors or televisions and pays recyclers to dispose of the displays safely when users no longer want them. Fees depend on the size of the monitor. In 2009, the fee to dispose of a monitor smaller than 15 inches was \$8, \$16 if the monitor is 15–35 inches, and \$25 for greater than 35 inches. In 2004,

TABLE 8.1 Weight and Gross Value of Selected Metals in Cell Phones in the United States

The average weight (wt) of a cell phone is estimated to be 113 grams (g), exclusive of batteries and charger (Nokia 2005). Metal contents are weights in metric tons (t), unless otherwise noted. Values in U.S. dollars are calculated by using the average of prices for 2002–2004 from USGS Mineral Commodity Summaries 2005 (Amey, 2005; Edelstein, 2005; Hilliard, 2005a, 2005b). The gross values do not include costs of recycling. Data may not add to totals shown because of independent rounding.

Metal	Metal Content and Value Estimated for a Typical Cell Phone		Metal Content and Value for 180 Million Cell Phones in Use in 2004[1]		Metal Content and Value for 130 Million Cell Phones Retired in 2005[1]		Metal Content and Value for 500 Million Obsolete Cell Phones in Storage in 2005[1]	
	Wt[2] (g)	Value ($)	Wt[3] (t)	Value ($ million)	Wt[3] (t)	Value ($ million)	Wt[3] (t)	Value ($ million)
Copper	16	0.03	2,900	6.2	2,100	4.6	7,900	17
Silver	0.35	0.06	64.1	11	46	7.9	178	31
Gold	0.034	0.40	6.2	72	3.9	52	17	199
Palladium	0.015	0.13	2.7	22.7	2.0	16	7.4	63
Platinum	0.00034	0.01	0.06	1.4	0.04	1	0.18	3.9
Total	2,973			113	2,152	82	8,102	314

[1]Number of cell phones in use in 2004 from Charny (2005). Number of cell phones retired in 2005 from U.S. Environmental Protection Agency, 2005. Number of obsolete cell phones projected to be in storage in 2005 from Most (2003).

[2]Metal content (wt) calculated from weight of a typical cell phone (Nokia 2005) and data from Rob Bouma, Falconbridge Ltd., written and oral communications, 2005.

[3]Metal content (wt) calculated from data from Rob Bouma, Falconbridge Ltd., written and oral communications, 2005.

Source: Daniel Sullivan. "Recycled Cell Phones—A Treasure Trove of Valuable Materials." USGS (2007), http://pubs.usgs.gov/fs/2006/3097/fs2006-3097.pdf.

TABLE 8.2 Current Electronics Recycling Laws in Effect and Year Passed

2003	California
2004	Maine
2005	Maryland
2006	Washington
2007	Connecticut, Minnesota, Oregon, Texas, North Carolina
2008	New Jersey, Oklahoma, Virginia, West Virginia, Missouri, Hawaii, Rhode Island, Illinois, and Michigan
2009	Indiana, Wisconsin
2010	Vermont, South Carolina, New York, Pennsylvania

Source: www.electronicsrecycling.org and www.epa.gov.

California passed a bill that makes it unlawful for retailers to sell mobile phones without the establishment of a collection, reuse, and recycling system for proper disposal of used cell phones. This bill places the responsibility for recycling squarely upon the industry, but leaves the implementation details up to them. While this approach allows the industry to minimize recycling costs, it remains to be seen whether the resulting policy promotes reuse of the materials in a manner that is safe for human health and the environment.

Other states have also focused on the manufacturer.[7] These states use market share as the basis for allocating responsibility for recycling to the manufacturers of televisions and video games. For example, for 2011 in the state of Maine, Samsung had a 19.6 percent share of recycling responsibility. Sony had a 11.3 percent share, Vizio 9.9 percent, etc., all the way to Audiovox with a 0.1 percent share.[8]

Are these laws working? Table 8.3 shows recycling rates for televisions, computer products, and cell phones in 2006–2007. While rates are up to 18 percent (from 15 through 2005), we still have a long way to go. It remains to be seen what effect the new laws and market share requirements will have.

Internationally, the Basel Convention regulates the movement of electronic waste across international boundaries (UNEP, 1989), although not all countries have ratified this treaty. One component of the convention would prohibit the export of e-waste from developed to industrializing countries since, in addition to valuable materials, the waste contains hazardous materials, such as lead and mercury.

In their analysis of trends of e-waste, Widner et al. (2005) find that for countries such as China and India, e-waste is rapidly growing from both domestic sources and illegal imports. These countries are just beginning to impose laws to control e-waste imports, but enforcement is lacking and the valuable materials create a business opportunity. Widner et al. estimate that 50–80 percent of collected domestic e-waste from nonratifying Basel Convention countries, such as the United States, is shipped to China and other Asian countries.

TABLE 8.3 Electronics Recycling Rates 2006–2007.

	Generated (million of units)	Disposed (million of units)	Recycled (million of units)	Recycling Rate (by weight)
Televisions	26.9	20.6	6.3	18%
Computer Products[1]	205.5	157.3	48.2	18%
Cell Phones	140.3	126.3	14.0	10%

[1]Computer products include CPUs, monitors, notebooks, keyboards, mice, and hard copy peripherals.

Source: Environmental Protection Agency (EPA), http://www.epa.gov/wastes/conserve/materials/ecycling/manage.htm.

[7]Details of each state's program can be found at http://www.electronicsrecycling.org/public/ContentPage.aspx?pageid=14

[8]http://www.maine.gov/dep/rwm/ewaste/manufacturers.htm

Pollution Damage

Another situation influences the use of recycled and virgin ores. When environmental damage results from extracting and using virgin materials and not from the use of recycled materials, the market allocation will be biased away from recycling. The damage might be experienced at the mine, such as the erosion and aesthetic costs of strip mining, or at the point of processing, where the ore is processed into a usable resource.

Suppose that the mining industry was forced to bear the cost of this environmental damage. What difference would the inclusion of this cost have on the scrap market? The internalizing of this cost results in a leftward shift in the supply curve for the virgin ore. This would, in turn, cause a leftward shift in the total supply curve. The market would be using less of the resource—due to higher price—while recycling more. Thus the correct treatment of these environmental costs would share with disposal costs a tendency to increase the role for recycling.

Disposal also imposes external environmental costs in the form of odors, pests, and contaminants leaching into water supplies; obstruction of visual landscapes; and so on. Kinnaman and Fullerton (2000) note that while the number of landfills in the United States has been decreasing, the aggregate capacity of these landfills has been increasing, as small-town facilities are replaced by large regional sanitary landfills. Since local opposition from potential host communities is likely to rise with landfill size, locating these facilities can be extremely contentious.

While governments now regulate landfills to protect public safety, these regulations rarely eliminate all unpleasant aspects of these landfills for the host communities. As a result, many communities are all for the existence of these facilities as long as they are not located in their community. If every community felt this way, locating new facilities could be difficult, if not impossible.

One technique for resolving this Not In My Back Yard (NIMBY) problem relies on the imposition of host fees. Host fees compensate the local community (and sometimes surrounding communities) for accepting the location of a waste facility within their community. This approach gives local communities veto power over the location, but it also attempts to share the benefits of the regional facility in such a way that makes the net benefits sufficiently positive for them that the communities will accept the facility.

In one example, Porter (2002) reports that a host fee agreement between Browning Ferris Industries and the township of Salem, Michigan, involves sharing with the town 2.5 percent of all landfill revenues and 4 percent of all compost revenues. The town also shares in the revenues derived from the sale of landfill gases (used for energy) and it can use the site free of charge for all town refuse, without limit on volume. These benefits are estimated to be worth about $400 per person per year, apparently enough to overcome local opposition.

Host fees are not a perfect resolution of the siting problem. Note, for example, that the fact that Salem can dispose of its waste free of charge provides no incentive for source reduction. In addition, it is important to ensure that locating these facilities does not raise environmental justice concerns. Although we consider this issue in much greater depth in Chapter 19, let it suffice here to point out that at a minimum, the local community has to be fully informed of the risks it will face from a regional sanitary landfill and must be fully empowered to accept or reject the proposed compensation package. As we shall see, these preconditions frequently did not exist in the past.

Additional complexities arise with hazardous wastes. Because hazardous wastes are more dangerous to handle and to dispose of, special polices have been designed to keep those dangers efficiently low. These policies will be also be treated in Chapter 19.

Summary

One of the most serious deficiencies in both our detection of scarcity and our ability to respond to scarcity is the failure of the market system to incorporate the various environmental costs of increasing resource use, be they radiation or toxics hazards, the loss of genetic diversity or aesthetics, polluted air and drinking water, or climate modification. Without including these costs, our detection indicators give falsely optimistic signals, and the market makes choices that put society inefficiently at risk.

As a result, while market mechanisms automatically create pressures for recycling and reuse that are generally in the right direction, they are not always of the correct intensity. Higher disposal costs and increasing scarcity of virgin materials do create a larger demand for recycling. This is already evident for a number of products, such as those containing copper or aluminum.

Yet a number of market imperfections tend to suggest that the degree of recycling we are currently experiencing is less than the efficient amount. The absence of sufficient stockpiles and the absence of tariffs mean that our national security interests are not being adequately considered. Artificially low disposal costs and tax breaks for ores combine to depress the role that old scrap can, and should, play. Severance taxes could provide a limited if poorly targeted redress for some minerals.

One cannot help but notice that many of these problems—such as pricing municipal disposal services and tax breaks for virgin ores—result from government actions. Therefore, it appears in this area that the appropriate role for government is selective disengagement complemented by some fine-tuning adjustments.

Disengagement is not the prescription, however, for environmental damage due to illegal disposal, air and water pollution, and strip mining. When a product is produced from virgin materials rather than from recycled or reusable materials, and the cost of any associated environmental damage is not internalized, some government action may be called for.

The selective disengagement of government in some areas must be complemented by the enforcement of programs to internalize the costs of environmental damage. The commonly heard ideological prescriptions suggesting that environmental problems can be solved either by ending government interference or by increasing the amount of government control are both simplistic. The efficient role for government in achieving a balance between the economic and environmental systems requires less control in some areas and more in others and the form of that control matters.

Discussion Questions

1. Glass bottles can be either recycled (crushed and re-melted) or reused. The market will tend to choose the cheapest path. What factors will tend to affect the relative cost of these options? Is the market likely to make the efficient choice? Are the "bottle bills" passed by many of the states requiring deposits on bottles a move toward efficiency? Why?
2. Many areas have attempted to increase the amount of recycled waste lubricating oil by requiring service stations to serve as collection centers or by instituting deposit-refund systems. On what grounds, if any, is government intervention called for? In terms of the effects on the waste lubrication oil market, what differences should be noticed among those states that do nothing, those that require all service stations to serve as collection centers, and those that implement deposit-refund systems? Why?
3. What are the income-distribution consequences of "fashion"? Can the need to be seen driving a new car by the rich be a boon to those with lower incomes who will ultimately purchase a better, lower-priced used car as a result?

Self-Test Exercises

1. Suppose a product can be produced using virgin ore at a marginal cost given by $MC_1 = 0.5q_1$ and with recycled materials at a marginal cost given by $MC_2 = 5 + 0.1q_2$. (a) If the inverse demand curve were given by $P = 10 - 0.5(q_1 + q_2)$, how many units of the product would be produced with virgin ore and how many units with recycled materials? (b) If the inverse demand curve were $P = 20 - 0.5(q_1 + q_2)$, what would your answer be?
2. When the government allows private firms to extract minerals offshore or on public lands, two common means of sharing in the profits are bonus bidding and production royalties. The former awards the right to extract to the highest bidder, while the second charges a per-ton royalty on each ton extracted. Bonus bids involve a single, up-front payment, while royalties are paid as long as minerals are being extracted.
   a. If the two approaches are designed to yield the same amount of revenue, will they have the same effect on the allocation of the mine over time? Why or why not?
   b. Would either or both be consistent with an efficient allocation? Why or why not?
   c. Suppose the size of the mineral deposit and the future path of prices are unknown. How do these two approaches allocate the risk between the mining company and the government?
3. "As society's cost of disposing of trash increases over time, recycling rates should automatically increase as well." Discuss.

4. Suppose a town concludes that it costs on average $30.00 per household to manage the disposal of the waste generated by households each year. It is debating two strategies for funding this cost: (1) requiring a sticker on every bag disposed of such that the total cost of the stickers for the average number of bags per household per year would be $30 or (2) including the $30 fee in each household's property taxes each year.
   a. Assuming no illegal disposal what approach would tend to be more efficient? Why?
   b. How would the possibility of rampant illegal disposal affect your answer? Would a deposit-refund on some large components of the trash help to reduce illegal disposal? Why or why not? What are the revenue implications to the town of establishing a deposit-refund system?

Further Reading

Dinan, Terry. "Economic Efficiency Aspects of Alternative Policies for Reducing Waste Disposal," *Journal of Environmental Economics and Management* Vol. 25 (1993): 242–256. Argues that a tax on virgin materials is not enough to produce efficiency; a subsidy on reuse is also required.

Jenkins, Robin R. *The Economics of Solid Waste Reduction: The Impact of User Fees* (Cheltenham, UK: Edward Elgar, 1993). An analysis that examines whether user fees do, in fact, encourage people to recycle waste; using evidence derived from nine U.S. communities, the author concludes that they do.

Kinnaman, T. C., and D. Fullerton. "The Economics of Residential Solid Waste Management," in T. Tietenberg and H. Folmer, eds. *The International Yearbook of Environmental and Resource Economics 2000/2001* (Cheltenham, UK: Edward Elgar, 2000): 100–147. A comprehensive survey of the economic literature devoted to household solid-waste collection and disposal.

Porter, Richard C. *The Economics of Waste* (Washington, DC: Resources for the Future, Inc., 2002). A highly readable, thorough treatment of how economic principles and policy instruments can be used to improve the management of a diverse range of both business and household waste.

Reschovsky, J. D., and S. E. Stone. "Market Incentives to Encourage Household Waste Recycling: Paying for What You Throw Away," *Journal of Policy Analysis and Management* Vol. 13 (1994): 120–139. An examination of ways to change the zero marginal cost of disposal characteristics of many current disposal programs.

Tilton, John E., ed. *Mineral Wealth and Economic Development* (Washington, DC: Resources for the Future, Inc., 1992). Explores why a number of mineral-exporting countries have seen their per capita incomes decline or their standards of living stagnate over the last several decades.

Additional References and Historically Significant References are available on this book's Companion Website: http://www.pearsonhighered.com/tietenberg/