5 *Dealing with Scale*

Executive Summary

- There is seldom one, ideal scale at which to conduct an ecosystem assessment that will suit several purposes. The Millennium Ecosystem Assessment (MA) advocates a multiscale approach.
- Many environmental problems originate from the mismatch between the scale at which ecological processes occur and the scale at which decisions on them are made. Outcomes at a given scale are often critically influenced by interactions of ecological, socioeconomic, and political factors from other scales. Focusing solely on a single scale is likely to miss such interactions, which are critically important in understanding ecosystem determinants and their implications for human well-being.
- The choice of scale and boundaries of an assessment is not politically neutral. It can implicitly favor certain groups, systems of knowledge, types of information, and modes of expression. Reflecting on the political consequences of scale and boundary choices is an important prerequisite to exploring how multiscale and cross-scale analysis in the MA might contribute to decision-making and public policy processes at various levels.
- Ecosystem processes and the services they deliver are typically most strongly expressed, most easily observed, or have their dominant drivers or consequences at particular scales in space and time. The spatial and temporal scales are often closely related, defining the scale domain of the process.
- Social, political, and economic processes can be more readily observed at some scales than others, and these may vary widely in terms of duration and extent. Furthermore, social organization has more- or less-discrete levels, such as the household, community, and nation, that correspond broadly to particular scale domains in time and space.
- Assessments need to be conducted within a scale domain appropriate to the processes or phenomena being examined. Those applicable to large areas generally use data at coarse resolutions, which may not detect fine-resolution processes. Even if data are collected at a fine level of detail, presentation of the findings at a larger scale means local patterns, anomalies, and the exceeding of thresholds disappear.
- A multiscale approach that simultaneously uses larger- and smaller-scale assessments can help identify important dynamics of the system that might otherwise be overlooked. Trends that occur at much larger scales, although expressed locally, may go unnoticed in purely local-scale assessments.

- If an assessment covers a shorter time period than the time scale of important processes, it will not adequately capture variability associated with, for instance, long-term cycles such as climatic or economic trends. Slow changes are often harder to detect than rapid changes, given the short period for which data are available.
- Given the pervasive influence of scale on any conclusions reached, it is essential that assessments be explicit regarding the geographic extent and period of time for which the study is valid. The same applies for data sets that are used in assessments.

Introduction

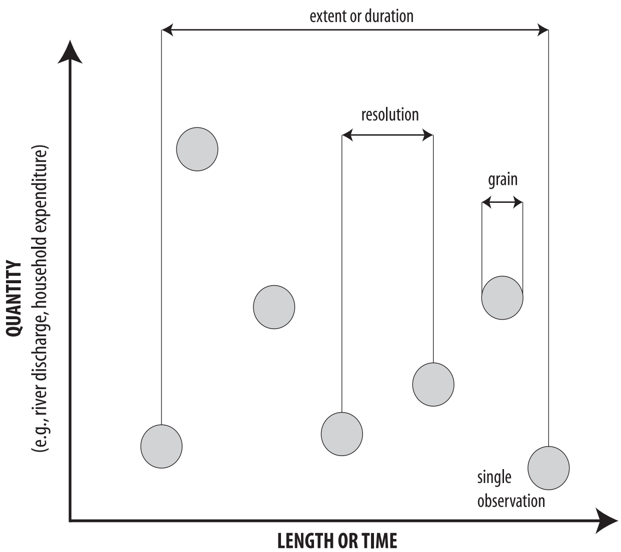
Scale refers to the physical dimensions, in either space or time, of phenomena or observations (O'Neill and King 1998). This is expressed in physical units, such as meters or years. In the Millennium Ecosystem Assessment (MA), the word "level" is used to describe the discrete levels of social organization, such as individuals, households, communities, or nations (Gibson et al. 2000). A level of organization is not a scale, but it can have a scale (Allen 1998; O'Neill and King 1998).

It is necessary to distinguish the "scale of observation" from the "scale of the phenomenon." The scale of observation is a construct based on human systems of measurement. Observation scale has three components: extent (or duration), resolution, and grain (Blöschl and Sivapalan 1995; Blöschl 1996). The extent is the total area or time over which a phenomenon is observed, the resolution is the interval or distance between observations, and the grain is the area or duration of an individual observation. These concepts are illustrated in Figure 5.1. Independent of the scales at which things are observed by humans and their instruments, there are characteristic scales at which both ecological and human processes occur. The characteristic scale of a process describes the typical extent or duration over which the process is expressed—that is, over which it has its impact. The scale domain of a process is defined in terms of both its characteristic space and time scales. The grain of a phenomenon is a concept distinct from the grain of observation, and refers to the smallest unit that is internally homogenous.

In the MA, unless otherwise stated, the word "scale" means the extent or duration of observation, analysis, or process. For instance, an assessment can be said to be "at the regional scale," or the time scale of the El Niño phenomenon is "at the decadal scale." The term "large scale" indicates something of greater extent than "small scale." This conforms to the

FIGURE 5.1 Three Components of Scale of Observation

The scale of observation can be described in terms of its extent, resolution, and grain. For example, in observing river discharge over time, grain refers to the time spent taking each sample, resolution refers to the time between observations, and extent refers to the total time period over which the samples were taken (based on Blöschl 1996). Similarly, in observing household expenditure in a particular area, grain refers to an individual household, resolution refers to the density and distribution of observed households over space, and extent refers the total area over which observations were made. In the special cases of continuous digital images or data recorders, grain is equal to resolution and in the former case is referred to as a pixel. Grain can also refer to the characteristics of the phenomenon itself—the smallest unit that is internally homogeneous, independent of the observer.



natural language usage of those phrases, although it is the opposite of common usage in cartography. "Long term" and "short term" are used in a relative sense for the time dimension, while "higher level" and "lower level" are used to reflect hierarchical institutional or social organizational levels. Higher levels have greater scope of operation or influence than lower levels.

Emergent properties may appear at some scales or levels of organization. The description of these properties is confined to those particular scales. An emergent property is a phenomenon that is not evident in the constituent parts of a system but appears when they interact as a system. For example, the cultural or recreational value of an ecosystem is often an emergent property at the scale of a landscape (a heterogeneous area consisting of a mixture of different patches, typically many kilometers in extent). There is debate about whether emergent properties have objective reality or are just a useful way of understanding and describing a system (Giampetro in press).

Scale is also related to variation and predictability: small-scale events show more variability than large-scale events do. This is because the effects of local heterogeneity are averaged out at broader scales, so that patterns appear to be more predictable (Weins 1989; Levin 1992). Conversely, models or assessments focusing on broad-scale patterns lose predictive accuracy at specific points in space and time (Costanza and Maxwell 1994).

Why Scale Matters

In assessing ecosystem services, scale matters for two main reasons. First, ecological and social systems and processes operate at a wide variety of scales—from very small and short to very large and long—and between scales they can change in their nature and sensitivity to various driving forces. It cannot be assumed that results obtained at one scale are automatically valid at another (Kremen et al. 2000; McConnell 2002).

Thus if the impacts of processes are observed or assessed at scales significantly smaller or larger than their characteristic scale, there is a likelihood of drawing the wrong conclusion. For instance, it is inappropriate to draw any conclusions regarding long-term trends based on short-duration time series data. People do not infer that the primary productivity of the world is declining just because in the Northern Hemisphere leaves die in the autumn; based on experience, it is obvious that this is part of a longer-term seasonal cycle. Nor can it be assumed that because a change is occurring at one location it is occurring equally at all locations. It is springtime in the Southern Hemisphere during autumn in the North.

Second, cross-scale interactions exert a crucial influence on outcomes at a given scale. Focusing solely on a single scale can miss these interactions. Looking at a particular issue top-down, from the perspective of larger scales or higher institutional levels, can lead to different conclusions than looking at the same issue bottom-up, from the perspective of smaller scales or lower levels (Berkes 2002; Lovell et al. 2002). The scale of the assessment influences both the framing of an issue and the range of possible actions and institutional responses. Where cross-scale interactions in ecological and social systems occur, there should be no expectation of finding a single most appropriate level for response or policy. In most cases, mutually supportive policy changes and responses at different levels are required in order to bring about desired results.

Individual systems important to ecological and social change arise from different scale domains of nature and society (Clark 1985; Peterson and Parker 1998). Scale has close connections with where, how, and by whom decisions are made regarding the use of ecosystem services. Scale also relates to how different actors learn about ecological change. Integrated understandings are of necessity nearly always place-based, and aggregates of observations drawn from smaller scales nearly always contain more detail than observations at a very general scale, offering opportunities for richer kinds of learning (e.g., Environment Canada 1997).

The "domain argument" (Wilbanks and Kates 1999) for conducting a multiscale assessment becomes even more persuasive when taking into account the many interactions across scales:

- Human rules and behavioral norms are embedded in scale-dependent institutional structures such as the boundaries of political jurisdictions. The spatial range of individual actions is restricted to the area of access rights: an agricultural plot, a forest patch, or a lake, for instance. Yet the social, economic, and political structures in which the actors are embedded are of larger scale—provincial, national, or even global. A local assessment is "local" not because it considers only local constraints and processes but because even while it takes account of factors and determinants from different scales, it is framed from the point of view of local stakeholders and it considers decisions and actions taken at that level. To be effective, local assessments must adequately reflect relevant factors and determinants from larger scales.
- Characteristic scales of ecological and human processes often do not match. Thus an integrated assessment of human-ecosystem interactions has to synthesize across scales (Rotmans and Rothman in press). For instance, the characteristic time scale for effective adaptation in ecosystem management depends on both human capabilities for changing management practices and the processes of structural change in the ecosystem.
- Cross-scale interactions can reveal hierarchical systems—that is, systems that "are analyzable into successive sets of subsystems" (Simon 1962:468). Hierarchical systems display a special type of nested orderliness and have particular resilience properties (Peterson 2000).

The minimum requirement for an assessment of ecosystem services is that it should be explicit about the scale and resolution of analysis. The

BOX 5.1 The Rationale for Multiscale Assessments

The Millennium Ecosystem Assessment is a multiscale assessment. The main reasons for introducing multiple scales in an already complicated assessment are the following:

- They permit individual ecological and social processes to be assessed at the scale at which they operate and to be linked to processes at different scales and levels of social organization.
- They allow progressively greater spatial, temporal, or causal detail to be considered as the scale becomes finer.
- They allow for independent validation of larger-scale conclusions by smaller-scale studies and create a context at larger scales for findings at smaller scales.
- They permit reporting and response options to match the scales at which social decision-making occurs, with which people can relate, and on which they can act—the local community, the province, the nation, the regional bloc, and the planet.

MA aspires to go beyond this by considering interactions across scales and levels of social organization. From the outset, the MA has been conceived as a multiscale effort, with attention to processes at several scales in space and time and at various institutional levels. (See Box 5.1.)

Changing Scales

Observations drawn from studies at widely different scales are only comparable with considerable care. Comparisons are valid only after careful testing to ensure that scale dependencies have been accounted for. Variables used to describe ecosystem services and their drivers can be thought of as belonging to one of three scaling categories: scale-independent, scale-dependent with known scaling rules, and non-scalable.

Scale-independent variables exhibit conservation of mass or value and show no (or weak) spatial or temporal interdependencies. To make the numerical values of such variables scale-independent, they can be divided by the measurement area (such as per square meter) or duration (per day, for instance). Population density (people per unit area) is an example. Scale-independent variables can be "scaled"—that is, translated from the scale at which the data were collected to a larger or smaller scale—in a very straightforward way through simple addition or proportionality. An example is biomass: the biomass of a hectare of forest is the simple sum of the biomass within each square meter of the forest. If the biomass is homogeneous over the scale of extrapolation, then every piece of the forest does not need to be measured to get the total. It is simply the total area multiplied by the biomass per unit area in the sub-sample.

In the second category, scale-dependent but with known (or potentially knowable) scaling rules, variables are "scalable"—that is, they can be expressed in smaller or larger aggregated units. But they must first be translated to a consistent scale, and the scaling rules may be complex and are often nonlinear. Transpiration is an example. Transpiration from a hectare of forest is not simply the transpiration measured at the scale of a leaf multiplied by the number of leaves in a hectare. This is because the transpiration from one leaf alters the humidity surrounding the leaves downwind of it, and thus their transpiration rate. Scaling of evapotranspiration can be achieved using an explicit model involving a nonlinear coupling constant (Jarvis and McNaughton 1986). Many social and ecological processes belong in this category. They tend to follow nonlinear or discontinuous scaling rules for a variety of reasons, including spatial or temporal interactions (especially feedbacks), organizational scope and the limits of institutional authority, and high heterogeneity or changes in the nature of the regulating factors as the scale changes.

Terrestrial carbon balance is an example of a variable that can be expressed in consistent physical units (grams of carbon per square meter per year) at all scales, but its interpretation changes with temporal and spatial scale. At the time scale of a few minutes and the scale of a leaf, the balance is called net photosynthesis (during the day) or respiration (at night). At the time scale of 24 hours or more, it is called net primary production (if considering plants only) or net ecosystem exchange (plants plus animals plus microbes). Over a period of decades or centuries, rare but large fluxes due to disturbance are included (such as fire, storms, harvest, or pest outbreaks), and the balance is called net biome production. The numerical value of net biome production is one hundredth or less the value of net photosynthesis.

Non-scalable variables or processes are those whose meaning is defined only at a particular scale. The process of decision-making within a household, for instance, may not be scaled up to the nation: different principles apply. Such variables can only be "qualitatively scaled" by placing them in clusters with conceptually related variables at different scales.

Assessments frequently compare or combine observations drawn from studies whose scales were determined independently. The main alternatives for doing so are to convert the observations to a single scale or to seek a multiscale or meta-scale synthesis. Conceptually, it has been suggested that convergence approaches (that is, bringing everything to a common scale) imply that process representations can be transferred seamlessly across scales, while multiscale approaches imply a rejection of that point of view (Bauer et al. 1999).

Converting scale-related information to a common metric often focuses on an intermediate scale, which calls for "downscaling" data about global processes and "upscaling" data about local processes (Wilbanks in press). Scaling up can be as intricate and fundamental a problem as downscaling. Harvey (2000) distinguishes between the use of lumped models, deterministically or statistically distributed models, and models with explicit spatial integration: lumped models use the same model representation for each scale, sometimes with implicit scaling in the form of parameter changes (Bugmann et al. 2000); distributed models use the same equation on a spatially explicit grid; and explicit integration seeks to formulate a correct representation for the higher-scale processes. Scaling up is essentially an aggregation challenge, complicated by the fact that simply adding smaller-scale values can give misleading results. For instance, the data may not meet standards for valid sampling, or they may fail to capture stochastic (random) variability in processes. The challenges are especially acute when larger-scale values are being estimated from incomplete local evidence. A number of technical alternatives for dealing with statistical problems in upscaling have been summarized (Rastetter et al. 1992; Harvey 1997).

Downscaling is a different challenge, involving collecting or estimating data at finer scales (such as regional or local values) from processes studied at larger scale. Modelers use both numerical (model-based) and empirical (data-based) approaches for this (Bass and Brook 1997; Easterling et al. 2000). Problems include limited data availability at detailed scales (and the costs of filling any gaps) and increasing complexity of causal relationships in more integrated, small-scale models. One of the forces encouraging downscaling is the need to provide information relevant to public participation, decision-making, and action at a relatively local scale.

Box 5.2 provides some practical guidance on dealing with scale-related data issues.

Space and Time Domains

One of the grand queries of science is understanding relationships between macro-scale and micro-scale phenomena and processes, between short-

BOX 5.2 Suggestions for Working with Information at Different Scales

If possible, convert studies of scale-dependent variables to a compatible scale before comparing or combining them. If not, interpret the studies independently at their individual scales.

Processes with nonlinear dynamics are seldom scale-independent:

- Weakly nonlinear processes may be approximated over limited parts of their range of scales by linear interpolation.
- For strongly nonlinear processes, a biased larger-scale estimate will be calculated if the inputs are averaged and then passed through the process. The correct approach is to calculate the output at each point for which input data are available, and then sum over space or time.

In upscaling, to generate an unbiased aggregation of a sparsely sampled variable from an uneven environment, use a richly sampled indicator that covaries with the variable of interest (a scalar) to create a weighted average.

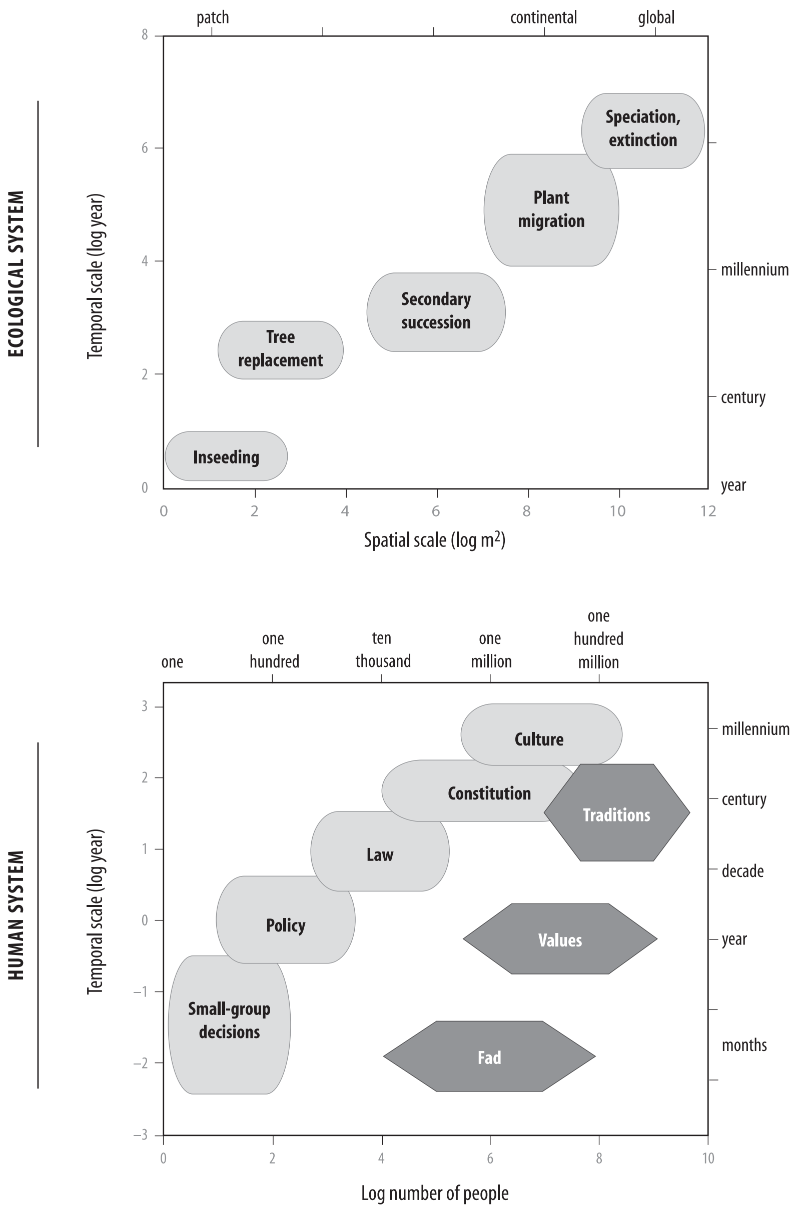
In downscaling, a probabilistic spatially or temporally explicit disaggregation of a heterogeneous variable can be constructed using a scalar.

It is not always necessary to drop to the finest resolution or to rise to the highest possible level of integration in order to represent cross-scale interactions adequately. To determine or illustrate the cause of a phenomenon, drop down to the next logical scale; to determine the constraints on a process, move up to the next logical scale. This is a rule of thumb, not a rigorous result, but is a useful way to limit the scope of an analysis.

term and long-term effects, and during implementation at different organizational levels (Wilbanks and Kates 1999; Gibson et al. 2000; Kates et al. 2003).

The time and space scales of a process are frequently correlated and are together referred to as the scale domain of the process (Bisonette 1997). “Big” processes are often “slow,” and “small” processes are often “fast.” (See Figure 5.2.) A fast process (or variable) is one that changes rapidly in relation to the life span of the organisms or entities that it is acting on. A slow process changes only gradually relative to the internal dynamics of the system being analyzed. In a forest ecosystem, for example, small and fast scales are dominated by biophysical processes controlling individual plant physiology and morphology. At the scale of a patch (tens of meters), interspecific competition for nutrients, light, and water influence growth, species composition, and succession over a period of decades. At forest stand scales, consisting of many patches, disturbances such as fire and insect outbreaks determine landscape heterogeneity over centuries. At the

FIGURE 5.2 Characteristic Scales in Time and Space for Some Ecological and Social Processes



Source: Modified from Delcourt et al. 1983; Gunderson et al. 1995a.

largest scales, climate processes alter structure and dynamics across hundreds of kilometers over thousands of years.

There are analogous space-time domains in social systems. For example, adaptive responses and behavioral changes at the individual level take place within an individual's lifetime, while societal responses often occurs over generations. In neither social nor ecological systems, however, is there necessarily a relationship between scales in space and time: some widespread changes have occurred very rapidly, and some local processes can be slow to change. The correlation between space and time scales can be particularly weak or absent in many modern social processes. It is speculated that this is due to the reach and speed of modern transport and information systems (Goodchild and Quattrochi 1997).

The correlation between large-scale ecosystem changes and long time frames presents a dilemma for ecosystem service assessments. To answer questions about the maintenance or resilience of these services in the future, long-term processes and their interaction with behavior on shorter time scales need to be understood. An implication is that global scale assessments, in particular, may need to consider historical and prehistorical data in order to gain the deeper time perspective necessary for a robust understanding of some large-scale processes. Furthermore, ecosystem assessments should strive to establish baselines against which future changes can be measured.

The links between characteristic scales in time and space can be used, cautiously, to infer long-term consequences by examining patterns over a large space domain. This is called a space for time substitution. For example, because large scales are likely to include areas undergoing rare events, instead of measuring net biome production over long periods (which is at any rate often impossible, given the short historical record), measurements can be done over large areas. Another example is the determination of long-term fire frequency. Many researchers have assumed that fire frequency is equal to the area fraction of the landscape burned per year. This will be untrue if it is the same parts of the landscape that burn repeatedly.

Inertia in Human and Ecological Systems

Both human and ecological systems frequently exhibit a property analogous to inertia in physical systems: the tendency to continue along a pathway of change after the pressure driving that change has been removed. The reason is that many of the processes involved have long time delays built into them. For instance, a fishery catch may continue to rise for a period after the point of sustainable catch has been exceeded, simply because of the maturation of juvenile fish that were hatched before the sustainability limit was passed (Rothschild 1986). Another example is sea level rise in response to climate change: this will continue for centuries after the emissions of greenhouse gases to the atmosphere have been radically reduced (IPCC 2002).

Inertia in the ecological system tends to mute the signals of impending problems and can lead to a tendency to overshoot the target once corrective action has been instituted. The inertia in human systems can cause the implementation of effective action to lag behind the first detection of the problem by years to decades. The combination of these two forms of inertia in coupled ecological-human systems has the potential to allow the system to transgress thresholds that are either ecologically unsustainable or socially unacceptable, and the resultant changes may be irreversible in realistic time frames.

There is a hypothesis, as yet unproven, that the "slow variables" (those with the largest inertia), rather than the "fast variables," are responsible for the resilience properties of a system (see papers in Gunderson and Holling 2002).

Viewing a Particular Scale in Context

Processes that operate at a particular scale are typically related to processes at other scales as well. One familiar example is land use at a local scale, which results from local institutions and actions but is shaped by national policy frameworks and global economic markets. At the same time, local actions may add up either as cumulative changes (such as species extinction) or as systemic changes at a larger scale (such as the effects of emissions of ozone-depleting gases on the stratosphere) (Turner II et al. 1990).

Since pieces of a geographic mosaic are nested within larger pieces, and those are within still larger pieces, it is often useful to think of geographic areas in terms of hierarchies of places and place-related processes. Such approaches lend themselves to the application of hierarchy theory (e.g., O'Neill 1988). Simon (1974) argued that semi-autonomous levels are formed from the interactions among a set of variables that share similar speeds and spatial domains.

In understanding social scales, it may be useful to consider the different forms of hierarchy. In inclusive hierarchies, groups of processes or objects lower in the hierarchy are contained within groups ranked as higher in the systems (such as modern taxonomic classifications). In exclusive hierarchies, groups of objects or processes that are ranked as lower are not contained within the groups at a higher level (military ranking systems, for example). And in a constitutive hierarchy, groups and processes are combined into new units with their own functions and emergent properties (a stakeholder committee, for instance).

Some important social processes do not neatly fit into this concept of nested hierarchies, with space and time strongly correlated. Social networks can introduce very strong connections between places, resulting in interaction across spatial and institutional boundaries. An example is the flow of ideas and coordination of action in different countries through transnational civil society organizations that do not necessarily flow through successive layers of a nested hierarchy. For instance, the Chipko (tree hugger) movement in India was a local-scale action that was quickly internationalized and that inspired similar social environmental movements worldwide. Such flows of ideas or advocacy for coordinated action also tend to be opportunistic, jumping over or skipping scales as they "shop" for the scale or forum that would provide the greatest chance of successful outcomes (Keck and Sikkink 1999).

Transfer of technologies and investments or sharing of stages of the commodity production chain through parts of a transnational corporation are other examples of crossing boundaries. Processes of diffusion of technological and institutional innovations are often critical factors in the use of ecosystem services. Thus network-related concepts are likely to be important in the MA for understanding connections between proximate and primary determinants or causal factors and for identifying possible response options. Network-related concepts are important as well in considering response options at small scales that may be replicated in larger domains without passing through the neat nested hierarchies of governance structures.

Scales in Ecological and Human Systems

The characteristic spatial scales of ecological systems are influenced by numerous factors, including the home range of individual mobile organisms or the range of influence of non-mobile organisms, the geographic distribution of a population of interbreeding organisms, the area over which a disturbance occurs, and the distance over which material is transported during the period for which it is ecologically active. For instance, the effective lifetime of carbon dioxide in the atmosphere is several centuries—during this time it can be transported all over the world. Hence its characteristic scale is global. In contrast, tropospheric ozone can only be transported by wind currents over a few hundred kilometers before it is consumed by atmospheric reactions; thus its characteristic scale is regional.

Characteristic temporal scales of ecological systems are influenced by the life span of organisms, the turnover rate of material pools, and the average period between disturbances at a location. An important distinction, particularly for determining system resilience, is between fast and slow variables or processes. Thresholds of irreversibility are typically related to changes in the slow variables (Gunderson and Holling 2002).

Spatial scales of social, political, and economic processes or variables are shaped by the area of operation, influence, or access rights exercised by various levels of institutions or social organization. Socioeconomic time scales are determined by the response times of humans and their institutions; they may be very rapid (electronic trading of commodities) or relatively slow (institutional change, typically). For example, the characteristic scale of an individual household in a freehold tenure system may be the area of land that the people own; for a community, it may be a village or municipal boundary; and for a country, it is the area included in the national borders and the exclusive economic zone in the ocean. The spatial scales of economic processes typically have a political dimension and are determined by the area over which goods or services are traded, extracted, transported, or disposed of. Economic and political processes are, in turn, embedded within and permeated by sociocultural processes that operate at different institutional levels.

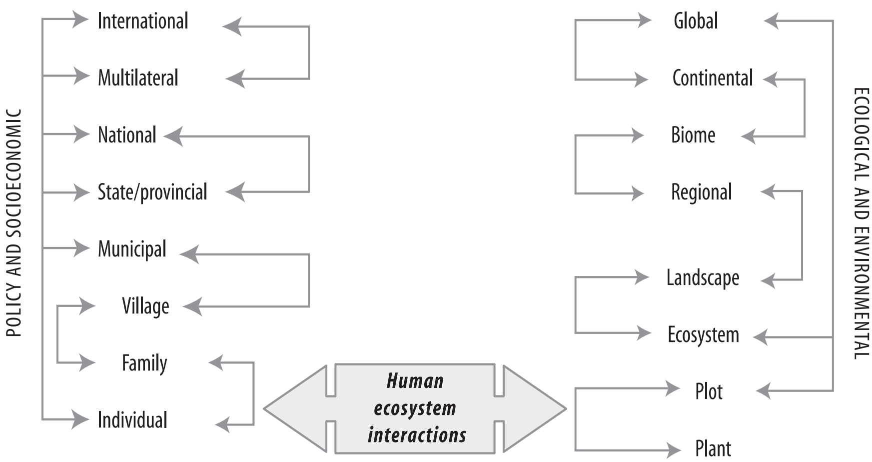
Direct interactions between humans and ecosystems—in agriculture, for example, or forestry or land use—mostly occur at local or micro scales and often at lower institutional levels. (See Figure 5.3.) This can also apply to indirect actions. For instance, although climate change is a global phenomenon, the responses of ecosystems are determined by the changes in the local climate rather than the global average change. Moreover, direct mitigation measures and behavioral responses also typically occur at the local level.

Ecosystem services, though often meeting needs expressed over large scales such as nations, are generally actually delivered at the local scale. Thus an assessment of ecosystem services and their implications for human well-being at global or regional scales typically needs to:

- scale up the ecosystem, taking each service in turn using specific scaling rules, including competition between different actors and different services; and

FIGURE 5.3 Overview of Some Commonly Used Institutional Levels and Ecological Scales

Levels are arranged on a shared vertical axis representing spatial extent. The arrows represent key influences. Direct interactions mostly take place at the local scale, but governance occurs at many scales.



Source: Courtesy of Rik Leemans.

- scale down impacts on ecosystems by scaling down either the environmental pressures (such as by regionalizing the estimates of global climate change) or the socioeconomic activities (by predicting, for example, where a logging company will harvest trees within a large concession area).

It is possible to assess ecosystem services and human well-being more readily at some spatial scales than others. For the MA, as an integrated assessment, these scales are determined by the characteristic scales of both ecological and socioeconomic processes. Thus the planet, the region (a fraction of Earth's surface corresponding to a biome or a major politico-economic bloc), the drainage basin, and the local community tend to emerge repeatedly as chosen scales for the MA.

Temporal scale is an issue in understanding human societies and institutions in a variety of ways. Perhaps the most familiar case is benefit-cost approaches to estimating the value of economic investments at different time scales. For example, in considering investment choices, how should long-term payoffs be evaluated relative to short-term payoffs? The conventional approach is to use a discount rate, which compares future returns from the investment with what could be earned from a neutral investment such as a mutual fund. Use of a high discount rate reduces the estimated value of longer-term returns, however, and is therefore biased in favor of investments that yield returns in the immediate or short term. (See also Chapter 6.)

Temporal scale is also a key consideration in historical studies, including studies of institutional, technological, and sociopolitical change. It is common to treat short-term, mid-term, and long-term forecasts differently in terms of methods and assumptions, because the longer term raises so many uncertainties about contextual assumptions. Quantitative forecasts of economic and demographic change are usually limited to time horizons of 25–30 years, if that; the longer term is considered the province of futurists rather than forecasters and is couched in terms of scenarios rather than predictions.

Most analyses of human systems include multiple time frames. For instance, national politics has a short-term rhythm set by the schedule of elections, while longer-term trends move in such directions as privatization, devolution, and democratization.

One subject of research has been whether long-term changes in human systems show regular, predictable fluctuations. As an example, there may be a Kondratiev cycle of 50–60 years in macroeconomies, perhaps related to waves of technological change, and Kuznets cycles of 15–20 years related to infrastructure development (Berry 1991). Recent studies point to possible fundamental relationships between geophysical rhythms and economic fluctuations (Berry 2000). Much of this work grows out of efforts to understand implications of natural climate variability over periods of millennia, but it also addresses issues of seasonal, annual, decadal, and century-scale environmental change.

Scale and Policy

Politics of Scale

The choice of scale is not politically neutral, because the selection may intentionally or unintentionally privilege certain groups. The adoption of a particular scale of assessment limits the types of problems that can be addressed, the modes of explanations that are allowed, and the generalizations that are likely to be used in analysis. This applies to temporal and spatial scales as well as institutional levels.

For example, the range of ecosystem services that are directly used and acknowledged as having important support functions is dependent on sociocultural contexts, and these are restricted in space. As the assessment is conducted at progressively larger scales, the number of ecosystem services that are fully shared among places, and thus can be mapped "wall-to-wall," drops. The local services that would be visible in a local assessment may no longer be visible in a regional or global assessment. The same basic ecosystem processes (such as net primary production) can be seen as providing different services at different scales—timber at the local scale, but carbon sequestration at the global scale. These issues are critical for the MA because trade-offs between the services are likely. At various scales we need to ask: ecosystem services for whom?

Analyzing these trade-offs requires understanding politics and markets. Many such cross-scale trade-offs are not perceived as such, but instead become conflicts or crises created by a more powerful group (often the state) around the provision of one favored form of ecosystem service from which they can obtain rent or other benefits. Scale can be an argument that empowers state institutions. Most states view indigenous knowledge and institutions as local in scope, relevance, and power, whereas the rules and knowledge of the state are viewed as bigger in scale, scope, and significance. As a consequence of this thinking, there is a strong tendency to override, minimize, or ignore local considerations, issues, or preferences. Many ecosystem management problems result from centralization and uniformity in bureaucratic operations that hinder local adaptation and learning. On the other hand, local adaptation is not universally good. Sometimes a state is needed to deal with the externalities that may arise from local decisions or to arbitrate among competing local claimants to ecosystem services. Scale is thus critical for issues of governance of ecosystems, as discussed in the next section.

Choice of time scales is equally important. If an assessment is focused on short-term concerns, then "important" ecosystem services are deemed to be those that are already or about to be threatened, such as freshwater resources for drinking, fuelwood supplies, or food production. On the other hand, if the users are more concerned with decisions that may have consequences over several decades or centuries, then issues of alterations of carbon balance or opportunity and resilience costs of biodiversity loss become much more important.

Adoption of a global scale immediately places issues such as climate change and carbon management at a much higher priority than, say, sanitation or access to clean drinking water. The attractiveness of the multiscale approach in the MA is that it provides the chance to think about problem identification and response options at more than one scale. It also allows for the analysis of spatial synergies and trade-offs among possible responses.

Likewise, the choice of boundaries is not neutral but has political consequences. For example, setting boundaries as a watershed basin as opposed to a geopolitical identity may make good sense from an ecological perspective, but it may be irrelevant for management if there is no political mechanism to deal with transboundary issues.

Reflection on the political consequences of scale and boundary choices is an important prerequisite to exploring what multiscale and cross-scale analysis in the MA might contribute to decision-making and public policy processes. Designating the boundaries is best done through collaboration between scientists, decision-makers, and representatives from different stakeholder groups.

Institutional Fit and Interplay

Many problems arise from the failure to recognize cross-scale interaction in both ecological and social systems (Young 1994). The effectiveness of institutions governing the management and use of ecosystem services depends not only on their own characteristics but also on how they interact with other institutions. An important class of interactions is those that occur vertically across levels of governance; these often correspond to changes in spatial scale.

The commonest forms of institutional interplay are those between state and local bodies. An emerging arena is the interplay between national institutions and new institutions at regional and international levels (Young 2002). These add a further layer of complexity to both future driving forces of change as well as possible sources of rules and guidance for human choice.

If cross-scale interactions in ecological and social systems affecting ecosystem services are common, then it should not be expected that there is generally a single most appropriate level for response or policy. While responses at certain levels or scales can have disproportionately greater significance or impact, appropriate responses at different levels are in general needed in concert to achieve desired results.

Guidance for Multiscale Assessments

Choosing the Appropriate Scales, Resolutions, and Boundaries

Several different approaches have been suggested for determining the most appropriate scale for an assessment (Wilbanks in press). One seeks the scale at which data show maximum inter-zonal variability and minimum intra-zonal variability. Another seeks the scale that minimizes statistical error between observed and modeled phenomena (Easterling et al. 1998). A third weighs increased information from finer spatial resolution against difficulties of gathering and analyzing the information (Costanza and Maxwell 1994).

The fact is, there is no single ideal scale for any instance of integrated assessment. The choice depends on the purposes of the analysis and is strongly conditioned by practical issues of data availability. The two most commonly used approaches are either to select a scale (often regional) on the basis of empirical evidence about the process involved (e.g., Kasperson et al. 1995; Schellnhuber and Wenzel 1998) or to select scales that correspond to human systems for decision-making (Cash and Moser 1998).

Despite the fact that the scale of a system is subjective (a function of the question being asked), the location of the boundaries should not be arbitrary. There are more and less useful places to locate the boundaries. The guiding principle is that a well-defined system has key feedbacks included in it and weak, slow, constant, or unidirectional interactions across the boundaries.

A practical approach to the spatial delimitation of an ecosystem is to build up a series of overlays of significant factors, mapping the location of discontinuities—for instance, in the distribution of organisms, the biophysical environment (soil types, drainage basins, shared markets), and spatial interactions (home ranges, migration patterns, fluxes of matter). A useful ecosystem boundary is one where a number of these relative discontinuities coincide. An ecosystem boundary can move over time. For instance, a marine ecosystem may be associated with an upwelling, which develops, moves, and dissipates. Similar approaches can be used to delimit human systems, such as the extent of particular patterns of land use or the political boundaries of a trade bloc. The systems addressed by the MA represent a pragmatic overlay of both ecosystem boundaries and human system boundaries.

Not all ecosystem services need to be addressed at every assessment scale. If there is a substantial mismatch between the characteristic scale of a process delivering a particular ecosystem service and the chosen scale of an assessment, it is preferable not to address that service then, leaving it to an assessment at a more appropriate scale (if such an evaluation exists).

These considerations of scale are the key reasons for performing a multiscale assessment. Comprehensive assessments need to be sensitive to multiple scales in time and space rather than focused on a single scale, and the local (or small-regional) end of the spectrum is often especially important. Processes at all relevant scales have to be included in the overall assessment, which requires appropriate methods to transfer, synthesize, and integrate information on data, variables, and processes between different scales. If results of smaller-scale studies are to be aggregated to larger scales, the prospects are brighter if they follow similar practices in the questions asked, the measurement or estimation approaches used, and the formats used for reporting results. Since time is unidirectional, temporal explicitness is typically inherent in the way observations are made and results reported. Studies must be equally explicit in the spatial dimension, although they seldom are.

The MA is designed as a collection of assessments, carried out partly independently at different scales, which are nested within one another in some cases. Effective approaches are needed for integrating top-down and bottom-up perspectives, particularly in the institutional domain, although the state of the art for such integration is not yet fully developed (Wilbanks in press).

Integration across Scales

Perhaps the greatest scale-related challenge to integrated, multiscale assessment is identifying, analyzing, and understanding linkages across scales. That they exist and are important is beyond question. The approach most often used is to analyze processes at several scales and then to examine how the findings at different levels correspond (e.g., Wu and Loucks 1995). An approach termed "strategic cyclical scaling" has been suggested in global change studies (Root and Schneider 1995). This calls for iterative cycling between upscaling and downscaling efforts, with each stage offering insights about the next as an understanding of cross-scale interactions grows.

Other suggested approaches tend to be more theoretical than practical. For instance, it is possible to think about interaction across scales as an extension of hierarchy theory (Allen and Starr 1982; O'Neill 1988). Hierarchies of scale-related processes define "constraint envelopes" within which subordinate elements of the hierarchy operate. Other possible approaches include system dynamics and dynamic spatial simulation modeling.

One relevant body of recent literature is associated with the work by the Resilience Alliance (Gunderson and Holling 2002). If systems are viewed at sufficiently long time scales, then the idea of an adaptive cycle (or configuration of states) may be valuable and can also be applied at various spatial scales.

6 Concepts of Ecosystem Value and Valuation Approaches

Executive Summary

- Decision-making concerning ecosystems and their services can be particularly challenging because different disciplines, philosophical views, and schools of thought conceive of the value of ecosystems differently.
- In the utilitarian (anthropocentric) concept of value, ecosystems and the services they provide have value to human societies because people derive utility from their use, either directly or indirectly (use values). People also value ecosystem services that they are not currently using (non-use values).
- Under the utilitarian approach, numerous methodologies have been developed to try to quantify the benefits of different ecosystem services. These are particularly well developed for provisioning services, but recent work has also improved the ability to value regulating, supporting, and cultural services. The choice of valuation technique is dictated by the characteristics of each case and by data availability.
- Non-utilitarian value proceeds from a variety of ethical, cultural, religious, and philosophical bases. These differ in the specific entities that are deemed to have value and in the interpretation of what having non-utilitarian value means. Notable among these are ecological, sociocultural, and intrinsic values. These may complement or counter-balance considerations of utilitarian value. The legal and social consequences for violating laws or regulations based on an entity's intrinsic value may be regarded as a measure of the degree of that value ascribed to them.
- The Millennium Ecosystem Assessment plans to use valuation as a tool that enhances the ability of decision-makers to evaluate trade-offs between alternative ecosystem management regimes and courses of social actions that alter the use of ecosystems and the services they provide. This usually requires assessing the change in the mix of services provided by an ecosystem resulting from a given change in its management.
- Most of the work involved in estimating the change in the value of ecosystem benefits concerns estimating the change in the physical flow of benefits (quantifying biophysical relations) and tracing through and quantifying a chain of causality between changes in ecosystem condition and human well-being. A common problem in valuation is that information is only available on some of the links in the chain, and often in incompatible units.

- Ecosystem values in terms of services provided are only one of the bases on which decisions on ecosystem management are and should be made. Many other factors, including notions of intrinsic value and other objectives that society might have, such as equity among different groups or generations, will also feed into the decision framework.

Introduction

The importance or "value" of ecosystems is viewed and expressed differently by different disciplines, cultural conceptions, philosophical views, and schools of thought (Goulder and Kennedy 1997). One important aim of the Millennium Ecosystem Assessment (MA) is to analyze and as much as possible quantify the importance of ecosystems to human well-being in order to make better decisions regarding the sustainable use and management of ecosystem services.

Understanding the impact of ecosystem management decisions on human well-being is an important objective. But if this information is presented solely as a list of consequences in physical terms—so much less provision of clean water, perhaps, and so much more production of crops—then the classic problem of comparing apples and oranges applies. The purpose of economic valuation is to make the disparate services provided by ecosystems comparable to each other, using a common metric. This is by no means simple, either conceptually or empirically. Society's ability to do so has increased substantially in recent years, however.

Ecosystems have value because they maintain life on Earth and the services needed to satisfy human material and nonmaterial needs. In addition, many people ascribe ecological, sociocultural, or intrinsic values to the existence of ecosystems and species. The MA recognizes these different paradigms, based on various motivations and concepts of value, along with the many valuation methods connected with them.

Ecosystems and the provisioning, regulating, cultural, and supporting services they provide have economic value to human societies because people derive utility from their actual or potential use, either directly or indirectly (known as use values). People also value ecosystem services they are not currently using (non-use values). This paradigm of value is known as the utilitarian (anthropocentric) concept and is based on the principles of humans' preference satisfaction (welfare).

Another set of values placed on ecosystems can be identified as the sociocultural perspective: people value elements in their environment based on different worldviews or conceptions of nature and society that are ethical, religious, cultural, and philosophical. These values are expressed through, for example, designation of sacred species or places, development of social rules concerning ecosystem use (for instance, "taboos"), and inspirational experiences. For many people, sociocultural identity is in part constituted by the ecosystems in which they live and on which they depend—these help determine not only how they live, but who they are. To some extent, this kind of value is captured in the concept of "cultural" ecosystem services. To the extent, however, that ecosystems are tied up with the very identity of a community, the sociocultural value of ecosystems transcends utilitarian preference satisfaction.

A different source of the value of ecosystems has been articulated by natural scientists in reference to causal relationships between parts of a system—for example, the value of a particular tree species to control erosion or the value of one species to the survival of another species or of an entire ecosystem (Farber et al. 2002). At a global scale, different ecosystems and their species play different roles in the maintenance of essential life support processes (such as energy conversion, biogeochemical cycling, and evolution). The magnitude of this ecological value is expressed through indicators such as species diversity, rarity, ecosystem integrity (health), and resilience. With increasing scarcity of space, and with limited financial resources, priorities have to be set regarding the conservation of the remaining biodiversity at all scale levels. The selection of protected areas and the determination of safe minimum standards regarding (sustainable) use of ecosystem services are based in part on these ecological values and criteria. The concept of ecological value is captured largely in the "supporting" aspect of the MA's definition of ecosystem services.

Although the various value paradigms have no common denominator and may lack any basis for comparison, some valuation approaches corresponding to them overlap and interact in various ways. Human preferences for all values can, to some extent, be measured with economic valuation methods, but ecological, sociocultural, and intrinsic value concepts have separate metrics and should be used in the decision-making process in their own right.

This chapter reviews the merits and deficiencies of these different valuation paradigms and how they complement or bound each other in assisting decisions and policy formulation for sustainable management and use of ecosystems. Ecological values are not discussed further here because they are dealt with extensively in Chapter 2.

The Utilitarian Approach and Economic Valuation Methods

The utilitarian paradigm of value is based on the fact that human beings derive utility from ecosystem services either directly or indirectly, whether currently or in the future. Two aspects of this paradigm need to be stressed. First, the use that an individual human being derives from a given ecosystem service depends on that individual's motivations, including, for example, his or her needs and personal preferences. The utilitarian approach, therefore, bases its notion of value on attempts to measure the specific usefulness that individual members of society derive from a given service, and then aggregates across all individuals, usually weighting them all equally.

Second, utility cannot be measured directly. In order to provide a common metric in which to express the benefits of the widely diverse variety of services provided by ecosystems, the utilitarian approach usually attempts to measure all services in monetary terms. This is purely a matter of convenience, however, in that it uses units that are well recognized, saves the effort of having to convert values already expressed in monetary terms into some other unit, and facilitates comparison with other activities that also contribute to well-being, such as spending on education or health. It explicitly does not mean that only services that generate monetary benefits are taken into consideration in the valuation process. On the contrary, the essence of practically all work on economic valuation of environmental and natural resources has been to find ways to measure benefits that do not enter markets and so have no directly observable monetary benefits.

Motivations for Economic Valuation

The most common reasons for undertaking a valuation of ecosystems are:

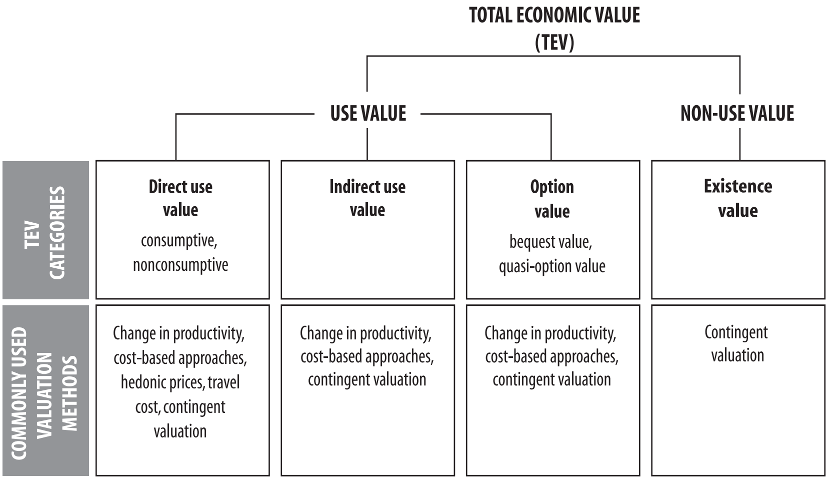
- to assess the overall contribution of ecosystems to social and economic well-being,
- to understand how and why economic actors use ecosystems as they do, and
- to assess the relative impact of alternative actions so as to help guide decision-making.

Numerous studies have assessed the contribution of ecosystems to social and economic well-being (Hartwick 1994; Asheim 1997; Costanza et al. 1997; Pimentel and Wilson 1997; Hamilton and Clemens 1999). Ecosystems form part of the total wealth of nations and contribute flow benefits, including social and cultural. But many ecosystem services are not traded, and hence their values are not captured in the conventional system of national accounts as part of total income. Moreover, in spite of the significant share of natural capital in total national wealth (World Bank 1997), the value of its depletion or appreciation is typically not accounted for.

As a result, conventional measures of wealth give incorrect indications of the state of well-being, leading to misinformed policy actions and ill-advised strategic social choices. For example, liquidation of natural assets to finance current consumption may appear to increase well-being when it does not take into account the corresponding decline in the capacity of the natural system to sustain the flow of economic, ecological, social, and cultural benefits in the future. More appropriate indicators that account for the flow and asset values of ecosystems are crucial for accurate monitoring of the implications of changes in ecosystem conditions for well-being. This is critical for the sustainable use and inter-temporal allocation of natural resources and for intergenerational equity. Valuation can help establish ecosystem values that allow correction of a country's national accounts (sometimes known as "greening") and construction of improved indicators of changes in wealth and well-being. Better valuation of the services provided by a given ecosystem does not guarantee that it will be conserved, as the costs of conservation might still be found to exceed its benefits, but it will almost certainly result in a lower loss of ecosystem services than otherwise.

Understanding why and how humans use ecosystems the way they do—for instance, why they cut natural forests, deplete soils, or pollute water surfaces—is a second reason to undertake a valuation of ecosystems. Markets guide the behavior and choices of individuals and public and private decisions. There is often a divergence, or wedge, between the market prices of goods and services as seen by individual economic agents and the social opportunity cost of using them. In particular, many services provided by ecosystems tend to be underpriced or not priced at all, leading to the inefficient and, often, unsustainable use of resources. By showing the existence and magnitude of differences between these private and social costs and benefits, valuation can help reveal policy and institutional failures (such as open access, public goods and externalities, or missing or incomplete markets), providing useful policy information on alternative intervention options for correcting them, such as creating markets or improving incentives.

FIGURE 6.1 The Total Economic Value Framework



The MA plans to use valuation primarily for the third rationale for undertaking it: assessing the impacts—the gains and losses—of alternative ecosystem management regimes. This provides a tool that enhances the ability of decision-makers to evaluate trade-offs between alternative ecosystem management regimes and courses of social actions that alter the use of ecosystems and the multiple services they provide.

It must be stressed that the ecosystem values in the sense discussed in this section are only one of the bases on which decisions on ecosystem management are and should be made. Many other factors, including notions of intrinsic value, as discussed later in this chapter, and other objectives that society might have, such as equity among different groups or generations, will also feed into the decision-making framework. (See Chapter 8.)

Total Economic Value

The concept of total economic value (TEV) is a widely used framework for looking at the utilitarian value of ecosystems (Pearce and Warford 1993). (See Figure 6.1.) This framework typically disaggregates TEV into two categories: use values and non-use values.

Use value refers to the value of ecosystem services that are used by humans for consumption or production purposes. It includes tangible and intangible services of ecosystems that are either currently used directly or indirectly or that have a potential to provide future use values. The TEV separates use values as follows:

- *Direct use values*. Some ecosystem services are directly used for consumptive (when the quantity of the good available for other users is reduced) or nonconsumptive purposes (no reduction in available quantity). Harvesting of food products, timber for fuel or construction, medicinal products, and hunting of animals for consumption from natural or managed ecosystems are all examples of consumptive use. Nonconsumptive uses of ecosystem services include enjoying recreational and cultural amenities such as wildlife and bird-watching, water sports, and spiritual and social utilities that do not require a harvesting of products. This category of benefits corresponds broadly to the MA description of provisioning and cultural services.
- *Indirect use values*. A wide range of ecosystem services are used as intermediate inputs for production of final goods and services to humans such as water, soil nutrients, and pollination and biological control services for food production. Other ecosystem services contribute indirectly to the enjoyment of other final consumption amenities, such as water purification, waste assimilation, and other regulation services leading to clean air and water supplies and thus reduced health risks. This category of benefits corresponds broadly to the MA notion of regulating and supporting services.
- *Option values*. Despite the fact that people may not currently be deriving any utility from them, many ecosystem services still hold value for preserving the option to use such services in the future either by the individual (option value) or by others or heirs (bequest value). Quasi-option value is a related kind of value: it represents the value of avoiding irreversible decisions until new information reveals whether certain ecosystem services have values that are currently unknown. (Note that some analysts place option value as a subset of non-use value rather than of use value, but they do not otherwise treat it differently.) This category of benefits includes provisioning, regulating, and cultural services to the extent that they are not used now but may be used in the future.

Non-use values are also usually known as existence value (or, sometimes, conservation value or passive use value). Humans ascribe value to knowing that a resource exists, even if they never use that resource directly. This is an area of partial overlap with the non-utilitarian sources of value discussed later in this chapter. The utilitarian paradigm itself has no notion of intrinsic value. However, many people do believe that ecosystems have intrinsic value. To the extent that they do, this would be partially reflected in the existence value they place on that ecosystem, and so would be included in an assessment of its total economic value under the utilitarian approach. This kind of value is the hardest, and the most controversial, to estimate.

Economic Valuation Methods

Under the utilitarian approach, numerous methodologies have been developed to attempt to quantify the benefits of different ecosystem services (Hufschmidt et al. 1983; Braden and Kolstad 1991; Hanemann 1992; Freeman III 1993; Dixon et al. 1994). As in the case of private market goods, a common feature of all methods of economic valuation of ecosystem services is that they are founded in the theoretical axioms and principles of welfare economics. These measures of welfare change are reflected in people's willingness to pay (WTP) or willingness to accept (WTA) compensation for changes in their level of use of a particular good or bundle of goods (Hanemann 1991; Shogren and Hayes 1997). Although WTP and WTA are often treated as interchangeable, there are important conceptual and empirical differences between them. Broadly speaking, WTP is appropriate when beneficiaries do not own the resource providing the service or when service levels are being increased, while WTA is appropriate when beneficiaries own the resource providing the service or when service levels are being reduced. In practice, WTA estimates tend to be substantially higher than WTP estimates. For this reason, WTP estimates are often used, as they are more conservative.

The methods commonly used to estimate the value of various services are shown in Figure 6.1. A number of factors and conditions determine the choice of measurement method. For instance, when an ecosystem service is privately owned and traded in the market, its users have the opportunity to reveal their preferences for such a good compared with other substitutes or complementary commodities through their actual market choices, given relative prices and other economic factors. For such ecosystem services, a demand curve can be directly specified based on observed market behavior. Many ecosystem services are not privately owned or traded, however, and hence their demand curves cannot be directly observed and measured. Alternative methods have been used to derive values in these cases. Different users and authors often classify the various methods of measuring ecosystem services values differently, but the grouping and naming systems converge to a broad classification that basically depends on whether the measures are based on observed or hypothetical behavior.

BOX 6.1 Valuation of Economic Services Through Observed Behavior

- *Direct observed behavior methods.* These methods derive estimates of value from the observed behavior of producers and consumers. They often use market prices and are most often applicable in cases where the ecosystem services are privately owned and traded in functioning markets. This approach is most frequently applicable to consumptive use, where goods are extracted from ecosystems and traded on markets.
- *Indirect observed behavior methods.* This category also uses actual observed behavior data but not on the ecosystem service in question. In the absence of actual market behavior regarding that particular service, these methods use observations on actual behavior in a surrogate market, which is hypothesized to have a direct relationship with the ecosystem service value. Examples in this category include hedonic pricing methods (which use statistical techniques to break down the price paid for a service into the implicit prices for each of its attributes, including environmental attributes such as access to recreation or clean air) and travel cost methods (which use observed costs to travel to a destination to derive demand functions for that destination). This group also includes cost-based methods (such as replacement cost methods, which value services at the cost of replacing, for example, a water purification service provided by an ecosystem with a new water treatment plant) that do not exactly reflect welfare (benefit-based) measures of value. (They sometimes underestimate and sometimes overestimate value.)

The standard valuation approach that uses actual observed behavior data is further divided into direct and indirect observed behavior methods. (See Box 6.1.) When they can be applied, these are generally considered preferable to measures based on hypothetical behavior.

The second valuation approach uses measures of economic value based on hypothetical behavior. In this category of methods, people's responses to direct questions describing hypothetical markets or situations are used to infer value. This group can also be divided into direct hypothetical (such as contingent valuation, in which respondents are asked directly how much they would be willing to pay for specified benefits) and indirect hypothetical measures of WTP or WTA (contingent ranking or conjoint valuation, which ask respondents to rank different bundles of goods).

A final category of approach is known as benefits transfer. This is not a methodology per se but rather the use of estimates obtained (by whatever method) in one context to estimate values in a different context. For example, an estimate of the benefit obtained by tourists viewing wildlife in one park might be used to estimate the benefit obtained from viewing wildlife in a different park. Benefits transfer has been the subject of considerable controversy in the economics literature, as it has often been used inappropriately. A consensus seems to be emerging that benefits transfer can provide valid and reliable estimates under certain conditions. These include that the commodity or service being valued is identical at the site where the estimates were made and the site where they are applied and that the populations affected have identical characteristics. Of course, the original estimates being transferred must themselves be reliable for any attempt at transfer to be meaningful.

Each of these approaches has seen broad use in recent years, and an extensive literature exists on their application. These techniques can and have been applied to a wide range of issues, including the valuation of cultural benefits (Pagiola 1996; Navrud and Ready 2002). In general, more direct measures are preferred to indirect ones. However, the choice of valuation technique in any given instance will be dictated by the characteristics of the case and by data availability.

Several techniques have been specifically developed to cater to the characteristics of particular problems. The travel cost method, for example, was developed to measure the utility derived by visitors to sites such as protected areas. The change in productivity approach, on the other hand, is quite broadly applicable to a wide range of issues. Contingent valuation is potentially applicable to any issue, simply by phrasing the questions appropriately, and as such has become widely used—probably excessively so, as it is easy to misapply and, being based on hypothetical behavior, is inherently less reliable. Data availability is a frequent constraint and often restricts the choice of approach. Hedonic price techniques, for instance, require vast amounts of data, thus limiting their applicability.

Putting Economic Valuation into Practice

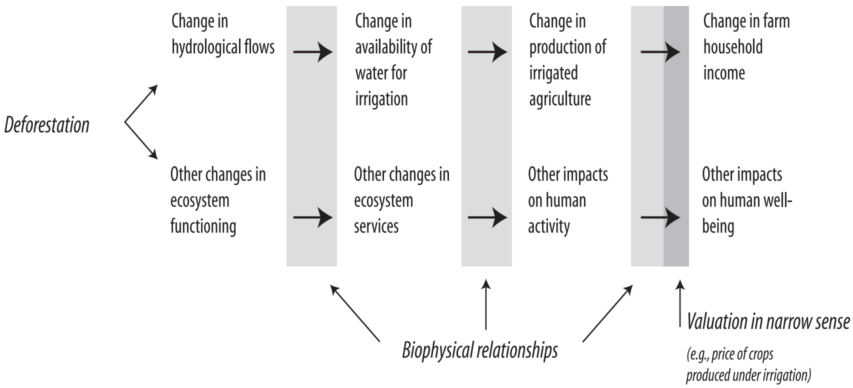
Whichever method is used for valuing a service, the analysis must begin by framing appropriately the question to be answered. In most policy-relevant cases, the concern is over changes in the level and mix of services provided by an ecosystem. At any given time, an ecosystem provides a specific "flow" of services, depending on the type of ecosystem, its condition (the "stock" of the resource), how it is managed, and its socioeconomic context. A change in management (whether negative, such as deforestation, or positive, such as an improvement in logging practices) will change the condition of the ecosystem and hence the flow of benefits it is capable of generating. It is rare for all ecosystem services to be lost entirely; a forested watershed that is logged and converted to agriculture, for example, may still provide a mix of provisioning, regulating, supporting, and cultural services, even though both the mix and the magnitude of specific services will have changed. Consequently, an assessment of the change in the value of services resulting from a given change in ecosystem management typically is most relevant to decision-makers and policy-makers. Where the change does involve the complete elimination of ecosystem services, such as the conversion of an ecosystem through urban expansion or road-building, then the change in value would equal the total economic value of the services provided by the ecosystem. (Measurements of total economic value of the services from a particular ecosystem can also be useful to policy-makers as an economic indicator, just as measures of gross national product or genuine savings provide policy-relevant information on the state of the economy.)

An assessment of the change in value of ecosystem services can be achieved either by explicitly estimating the change in value or by separately estimating the value of ecosystem services under the current and the alternative management regime and then comparing them. If the loss of a given service is irreversible, then the loss of the option value of that service will also be included. (An important caveat here is that the appropriate comparison is between the ecosystem with and without the management change; this is not the same as a comparison of the ecosystem before and after the management change, as many other factors will usually also have changed.) The typical question being asked, then, is whether the total value of the mix of services provided by an ecosystem managed in one way is greater or smaller then the total value of the mix provided by that ecosystem managed in another way.

The actual change in the value of the benefits can be expressed either as a change in the value of the annual flow of benefits, if these flows are relatively constant, or as a change in the present value of all future flows. The latter is equivalent to the change in the capital value of the ecosystem and is particularly useful when future flows are likely to vary substantially over time. (It is important to bear in mind that the capital value of the ecosystem is not separate and additional to the value of the flows of benefits it generates; rather, the two are intimately linked in that the capital value is the present value of all future flows of benefits.)

Estimating the change in the value of the flow of benefits provided by an ecosystem begins by estimating the change in the physical flow of benefits. This is illustrated in Figure 6.2 for a hypothetical case of deforestation that affects the water services provided by a forest ecosystem.

FIGURE 6.2 Valuing the Impact of Ecosystem Change



Source: Adapted from Pagiola et al. in press.

The bulk of the work involved in the exercise actually concerns quantifying the biophysical relationships. In many cases, this requires tracing through and quantifying a chain of causality. Thus, valuing the change in production of irrigated agriculture resulting from deforestation requires estimating the impact of deforestation on hydrological flows, determining how changes in water flows affect the availability of water to irrigation, and then estimating how changes in water availability affects agricultural production. Only at the end of this chain does valuation in the strict sense occur—when putting a value on the change in agricultural production, which in this instance is likely to be quite simple, as it is based on observed prices of crops and agricultural inputs. The change in value resulting from deforestation then requires summing across all the impacts.

Clearly, following through a chain like this requires close collaboration between experts in different disciplines—in this example, between foresters, hydrologists, water engineers, and agronomists as well as economists. It is a common problem in valuation that information is only available on some of the links in the chain, and often in incompatible units. The MA can make a major contribution by helping the various disciplines involved to become more aware of what is needed to ensure that their work can be combined with that of others to allow a full analysis of such problems.

In bringing the various strands of the analysis together, there are many possible pitfalls to be wary of. Inevitably, some types of value will prove impossible to estimate using any of the available techniques, either because of lack of data or because of the difficulty of extracting the desired information from them. To this extent, estimates of value will be underestimates. Conversely, there is an opposite danger that benefits (even if accurately measured) might be double-counted.

As needed, the analysis can be carried out either from the perspective of society as a whole ("social" analysis) or from that of individual groups within society ("private" analysis). Focusing on a particular group usually requires focusing on a subset of the benefits provided by an ecosystem, as that group may receive some benefits but not others. (Groups located within an ecosystem, for example, typically receive most of the direct use benefits but few of the indirect use benefits, whereas the opposite applies to downstream users.) It will often also require using estimates of value specific to that group; the value of additional water, for example, will be different depending on whether it is used for human consumption or for irrigation. The analysis can thus allow distributional impacts and equity considerations to be taken into account, as well as overall welfare impacts on society as a whole. This type of disaggregation is also useful in understanding the incentives that particular groups face in making their ecosystem management decisions. Many ecosystems are mismanaged, from a social perspective, precisely because most groups that make decisions about management perceive only a subset of the benefits the ecosystem provides.

Similarly, estimating the impact of changes in management on future flows of benefits allows for intergenerational considerations to be taken into account. Here, too, the bulk of the work involved concerns predicting the change in future physical flows; the actual valuation in the narrow sense forms only a small part of the work. Predicting the value that future generations will place on a given service is obviously difficult. Technical, cultural, or other changes could result in the value currently placed on a service either increasing or decreasing. Often, the best that can be done is to simply assume that current values will remain unchanged. If trends suggest that a particular change in values will occur, that can be easily included in the analysis. Such predictions are notoriously unreliable, however.

Non-utilitarian Value

From the perspective of many ethical, religious, and cultural points of view, ecosystems are valued even if they do not contribute directly to human well-being. Some ecosystems may be vital to a people's identity as a distinct society or culture. Thus preserving the health of such ecosystems may be a necessary condition for measuring changes in the collective welfare of those societies and cultures. Further, to the extent that a society's or a culture's ecocentric philosophical and ethical views recognize the intrinsic value of nonhuman species and ecosystems, sociocultural value also reaches beyond human welfare considerations.

Sociocultural Values

For many people, ecosystems are closely associated with deeply held historical, national, ethical, religious, and spiritual values. A particular mountain, forest, or watershed may, for example, have been the site of an important event in their past, the home or shrine of a deity, the place of a moment of moral transformation, or the embodiment of national ideals. These are some of the kind of values that the MA recognizes as the cultural services of ecosystems. And to some extent they are captured by utilitarian methods of valuation. But to the extent that some ecosystems are essential to a peoples' very identity, they are not fully captured by such techniques.

These values fall between the utilitarian and intrinsic value paradigms. They might be elicited by using, for example, techniques of participatory assessment (Campell and Luckert 2002) or group valuation (Jacobs 1997; Wilson and Howarth 2002). This evolving set of techniques is founded on the assumption that the valuation of ecological goods and services should result from a process of open public deliberation, not from the aggregation of separately measured individual preferences. Using this approach, small groups of citizens are brought together in a moderated forum to deliberate about the economic value of ecosystem goods or services (Wilson and Howarth 2002). The end result is a deliberative or "group" contingent valuation (CV) process (Jacobs 1997; Sagoff 1998). With a group CV, the explicit goal is to derive an economic value for the ecological good or service in question. The valuation exercise is conducted in a manner very similar to a conventional CV survey—using hypothetical scenarios and payment vehicles—with the key difference being that value elicitation is not done through private questioning but through group discussion and consensus building.

The Intrinsic Value Paradigm

Although the notion that nature has intrinsic value is a familiar one in many religions and cultures, it is unfamiliar in the context of modern rational choice theory and economic valuation. Yet analysts do have a well-established and familiar metric for assessing the intrinsic value of human beings and their various aspects. This valuation method and its metric may then be extended to some nonhuman natural entities, including ecosystems.

The notion that ecosystems have intrinsic value is based on a variety of points of view. Intrinsic value is a basic and general concept that is founded upon many and diverse cultural and religious worldviews. Among these are indigenous North and South American, African, and Australian cultural worldviews, as well as the major religious traditions of Europe, the Middle East, and Asia.

In the Judeo-Christian-Islamic tradition of religions, human beings are alleged to be created in the image of God. On that basis, humans are attributed intrinsic value. The Bible also represents God as having created plant and animal species, and declares the things thus created to be "good." Some commentators have argued that in doing so, God attributes intrinsic value to them, and thus that plant and animal species and the other aspects of nature that God also declared to be good have intrinsic value by an act of divine fiat (Barr 1972; Zaidi 1981; Ehrenfeld and Bently 1985).

In some American Indian cultural worldviews, animals, plants, and other aspects of nature are conceived as relatives, born of one universal Mother Earth and Father Sky (Hughes 1983). Thus they have the same value as human relatives: intrinsic value—if not in name, then at least in pragmatic effect. You may not sell your mother at any price; even performing a hypothetical economic valuation of your mother is questionable. And so, some American Indian elders have argued, neither should humans sell Mother Earth—that is, their tribal lands—or even compromise the intrinsic value of Earth by carrying out an economic valuation of tribal lands (Gill 1987).

Examples of other religious worldviews supporting the concept of intrinsic value in nature abound. Basic to Hindu religious belief is the essential oneness of all being, Brahman, which lies at the core of all natural things. The presence of Brahman in all natural things is the Hindu basis of intrinsic value (Deutch 1970). Closely related to this idea is the moral imperative of *ahimsa,* non-injury, extended to all living beings. The concept of *ahimsa* is also central to the Jain environmental ethic (Chapple 1986). Buddhism incorporates *ahimsa* as a central moral imperative as well (Chapple 1986). Also central to Buddhism is the overcoming of suffering by the cessation of desire. Absent desire, the natural world ceases to be referenced to a person as a pool of resources existing to satisfy desires or preferences (Kalupahana 1985). The enlightened Buddhist is thus able to appreciate the intrinsic value of nature.

Taoism, a major philosophical and religious tradition of China, posits the *Tao* or Way of nature as a norm of human action (Tu 1985). Taoism regards human economies as a subset of the economy of nature. In the Japanese Shinto religious tradition, the *kami* (gods), are closely associated with various aspects of nature (Odin 1991). As the *kami* have a greater-than-human dignity, the aspects of nature with which they are associated are also thought to have intrinsic value. In the Dreamtime narratives of the peoples indigenous to Australia, various features of the landscape are the places where the totemic Ancestors performed "terraforming" deeds (Stanner 1979). Such places are sacred and, in effect, have intrinsic value.

These are but a few of the bases for intrinsic value in non-western religious and cultural worldviews (for a comprehensive summary, see Callicott 1994). It is important for decision-makers to assess empirically the actual ecosystem-oriented values—intrinsic, sociocultural, and ecological, as well as utilitarian—of those affected by ecosystem-oriented policy and decisions.

The two main traditions of modern secular ethics in western culture are utilitarianism and Kantianism. In classical utilitarianism, aggregate "happiness," understood as a greater balance of pleasure over pain, was the putative goal of social policy. Contemporary economics is derived from utilitarianism and posits "preference satisfaction" as the goal of rational choice (Sen 1987). If aggregate preference satisfaction is, correspondingly, the goal of social policy, this may sometimes be maximized at the cost of overriding the interests of a comparatively few individuals (Rawls 1971). The potential injustices of unbridled utilitarianism are checked by the assertion of individual rights—most basically to life, liberty, and property.

Economic valuation of ecosystem services has been variously criticized by different commentators (e.g., Bromley 1990; Costanza 2000; Heal 2000a; Heal 2000b; Ludwig 2000; Pritchard et al. 2000). Further, reducing all values to preferences has been contested (Sagoff 1988). A person may prefer chocolate to vanilla ice cream, but some find it demeaning to the intrinsic value of human life and human liberty to say that as a society humans collectively prefer not to stage gladiator shows or own slaves or that, as an individual, a person merely prefers honesty over perfidy or justice over treachery.

The counter-utilitarian idea that there is a difference between preferences and values and that considerations of individual rights tempers calculations of aggregate utility was most clearly and powerfully expressed by Kant, who wrote, "Everything has either a *price* or a *dignity*. Whatever has a price can be replaced by something else as its equivalent; on the other hand, whatever is above all price, and therefore admits of no equivalent, has a dignity. But that which constitutes the condition under which alone something can be an end in itself does not have mere relative worth, i.e., a price, but an intrinsic worth, i.e., a dignity" (Kant 1959 [1785]:53, italics in original).

Because human rights, based on the dignity and intrinsic value of human beings, has traditionally been used to check the excesses and potential injustices of calculations of aggregate utility, many non-anthropocentric ethical theorists have largely adopted the intrinsic value paradigm. They first extended it to cover various nonhuman animals (Regan 1983). Some have attempted to push this line of argument further, to argue that all organisms have interests, goods of their own, natural goals, developments, and fulfillments and so should be accorded intrinsic value (Taylor 1986). Based on the seminal work of Aldo Leopold (1949), others have argued that transorganismic levels of biological organization (species, biotic communities, ecosystems) also have intrinsic value (Callicott 1989; Rolston III 1994). On whatever basis, intrinsic value has been attributed to various aspects of nature (genes, organisms, populations, species, evolutionarily significant units, biotic communities, ecosystems) and to nature as a whole (the biosphere).

The basis on which intrinsic value is attributed to various entities may limit which ones can have intrinsic value. For example, if being rational is the property required for something to have intrinsic value, then only rational beings (effectively, only human beings) are recognized to be intrinsically valuable. Non-anthropocentric theorists who have posited the criterion of "having interests" for ascribing intrinsic value thus limit it to individual organisms. In traditional Judeo-Christian thinking, those who thought that intrinsic value should be based on the property of being created in the image of God also effectively limit intrinsic value to human beings. In the Dreamtime worldview of the peoples indigenous to Australia, although landscape-level features have intrinsic value, individual plants and animals usually do not (except those associated with a person's own totem). Aldo Leopold (1949) thought that the things deserving of human "love and respect" had intrinsic value. Theoretically someone can love and respect anything at all, but Leopold argued that among other things, "biotic communities" commended themselves to human capacity for love and respect.

The Interactions of Political and Market Metrics

Parallel to using the market or its surrogates to measure economic value, in democratic societies the modern social domain for the ascription of intrinsic value is the parliament or legislature (Sagoff 1988). In other societies a sovereign power ascribes intrinsic value, although this may less accurately reflect the actual values of citizens than parliamentary or legislative acts and regulations do. The metric for assessing intrinsic value is the severity of the social and legal consequences for violating laws prohibiting a market in or otherwise compromising that which is recognized to be intrinsically valuable. In western societies long influenced by the Judeo-Christian worldview and Kantian moral philosophy, the highest intrinsic value is attributed to human life. Thus the severest of consequences are prescribed for murdering human beings.

Each kind of value—utilitarian, ecological, sociocultural, and intrinsic—is played out on a common and not always level playing field. Thus the various kinds of value intersect and interact in various ways. One common effect of socially recognizing and legally institutionalizing something's intrinsic value is to take it off the open market, to insist that it has a dignity and therefore should have no price. The clearest and most obvious example is human beings themselves. In most modern societies, there is no legal market in human beings; there is no open slave market. With the advent of human organ transplants, some societies have decided that there should be no legal market in human organs either; these are, by implication, thus accorded intrinsic value.

A black market often emerges in entities that are sufficiently well recognized as having a dignity to register a signal in the political intrinsic value metric. Depending on the strength of that signal—for instance, the social and legal consequences of pricing and trafficking in that entity—the supply of such entities declines and the price rises. So one effect of the political intrinsic value metric on the market metric is analogous to the effect of an excise tax or tariff.

Some things may arguably have both a dignity and a price—human labor, for example. Society may protect the recognized intrinsic value of things that also have utility by assuring, among other things, that their price is right. This may be the ethical rationale for minimum-wage laws, legally mandated health insurance, and retirement benefits in societies that have provided such protections by law. Society may also constrain the use of human labor with regulations designed to protect workers' health and safety.

Laws and regulations recognizing the intrinsic value of such things as endangered species, biodiversity more generally, and ecosystems such as wetlands have created a regulatory environment to which market forces are beginning to respond. A legal market in conservation "credits" is emerging. The red-cockaded woodpecker, for example, is a "listed" species protected by the U.S. Endangered Species Act (ESA), administered by the U.S. Fish and Wildlife Service (FWS). An agreement between International Paper (IP) and the FWS permits the company to consolidate at one location the breeding pairs of red-cockaded woodpeckers on its properties in several southeastern states and intensively manage that location as habitat for the endangered species. The agreement permits IP to harvest timber on the vacated sites and to sell credits to other owners of red-cockaded woodpecker habitat as the species recovers and the number of breeding pairs increases beyond a specified threshold (U.S. Fish and Wildlife Service 1999). Similarly, a company wishing to convert a wetland to a shopping mall faces regulatory constraints prohibiting wetland destruction. It can comply with those constraints by purchasing a credit from a distant landowner whose property contains a comparable wetland that will be protected (Fernandez 1999). This provides a market incentive to wetlands owners to conserve them.

Another effect of the political intrinsic value metric on the market metric is to shift the burden of proof away from those who would protect something with socially recognized and legally sanctioned intrinsic value and toward those who would commercially exploit it. The debate about human embryonic stem-cell research in the United States is a case in point. As aspects of human being, human embryonic stem cells are alleged to have a dignity and therefore should not be commercially exploited by the pharmaceutical industry, some have argued (with ambiguous political success). To overcome this argument, the pharmaceutical industry and its scientific allies must successfully counterargue that the aggregate utility of human embryonic stem-cell research is so great as to warrant overriding the putative dignity of this aspect of human being (Orkin and Morrison 2002).

Just because something has publicly recognized intrinsic value does not mean that its value is absolute or inviolable. Even human beings can be "converted" in deference to other values. Soldiers, for example, are often placed in harm's way to advance a country's perceived national interests or even aggregate economic welfare. In such cases, the intrinsic value of human beings seems sacrificed in favor of other values. But when intrinsic values are in zero-sum conflict with utilitarian values, the burden of proof rests with those advocating the latter.

Perhaps the most interesting and relevant case in point of legislative ascription of intrinsic value to some aspect of nature—and of the meeting of utilitarian and intrinsic value metrics—is the U.S. Endangered Species

Act enacted in 1973. In giving absolute legal protection to listed endangered species, the ESA, in effect, gave them a dignity comparable in strength to the dignity accorded individual human life. As noted, even the dignity of human life can be legally overridden, but the burden of proof falls on those who would do so. The ESA was amended in 1978 to create a Cabinet-level Endangered Species Committee empowered to decide whether opportunity cost (measured on the market metric) of protecting a listed species was high enough to warrant overriding its dignity (measured on the political metric).

This interaction between the political metric of intrinsic value and the market metric (and its surrogates) of utilitarian value has an analog in economic valuation called the safe minimum standard (SMS). Approaching the task of economically valuing ecosystem services by means of the SMS is practically equivalent to socially recognizing their intrinsic value and protecting them by law. Whereas benefit-cost analysis approaches each case and builds up a body of evidence about the benefits and costs of preservation, the SMS approach starts with a presumption that the maintenance of the healthy functioning of any ecosystem is a positive good (lumping together economic, ecological, sociocultural, and intrinsic values). The empirical economic question is, How high are the opportunity costs of satisfying the SMS? The SMS decision rule is to maintain the ecosystem unless the opportunity costs of doing so are intolerably high. The burden of proof is thus assigned to the case against maintaining the SMS (Randall 1998).

The quantitative threshold to which the opportunity costs must rise to warrant violating the SMS is left as an open empirical question. In practice, such thresholds are set by the political metric. The economic threshold for violating the SMS for ecosystem health will depend in part on how successful its advocates are in persuading voters that ecosystems have a dignity—not necessarily instead of, but as well as a price—and should be protected unless the opportunity costs of doing so are intolerably high. The question of how high is high enough will be indicated in part by the strength of laws and regulations enacted to protect ecosystems. In this case, however, the intrinsic value (assessed on the political metric) is augmented by the considerable utilitarian value of ecosystem services; their psycho-spiritual utilitarian values; their option, bequest, and existence utilitarian values; and their ecological and sociocultural values.

Conclusion

Human societies face important choices in how they manage ecosystems, affecting their conditions and the services they provide and thus ultimately human well-being. How decisions are made will depend on the systems of value endorsed in each society, the conceptual tools and methods at their disposal, and the information available. Making the appropriate choices requires, among other things, reliable information on actual conditions and trends of ecosystems and on the economic, political, social, and cultural consequences of alternative courses of action.

The MA will provide decision-makers with relevant information to aid them in making appropriate ecosystem management decisions. The impact that these decisions will have on human well-being is of particular interest. In some cases, these impacts can be assessed with indicators, such as the impact on human health. When there are multiple impacts and well-being is affected in many different ways, however, such unidimensional indicators will not be sufficient. In these cases, economic valuation will provide an important tool, as it will allow for different impacts to be compared and aggregated.

Of course, the importance of ecosystems goes beyond their role for human well-being. Non-utilitarian sources of value must also be taken into consideration in order to make appropriate management decisions.