Economic and ecological concepts for valuing ecosystem services

Stephen C. Farber a,*, Robert Costanza b,1,3, Matthew A. Wilson c,2,3

a Graduate School of Public and International Affairs, University of Pittsburgh, Pittsburgh, PA 15260, USA
b Center for Environmental Science and Department of Biology, Institute for Ecological Economics, University of Maryland, Box 38, 1 Williams St., Solomons, MD 20688-0038, USA
c Institute For Ecological Economics, University of Maryland, 0216 Symons Hall, College Park, MD 20742-5585, USA

Abstract

The purpose of this special issue is to elucidate concepts of value and methods of valuation that will assist in guiding human decisions vis-à-vis ecosystems. The concept of ecosystem service value can be a useful guide when distinguishing and measuring where trade-offs between society and the rest of nature are possible and where they can be made to enhance human welfare in a sustainable manner. While win-win opportunities for human activities within the environment may exist, they also appear to be increasingly scarce in a ‘full’ global ecological–economic system. This makes valuation all the more essential for guiding future human activity. This paper provides some history, background, and context for many of the issues addressed by the remaining papers in this special issue. Its purpose is to place both economic and ecological meanings of value, and their respective valuation methods, in a comparative context, highlighting strengths, weaknesses and addressing questions that arise from their integration. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Economic valuation; ecological valuation; ecological services; valuation

1. Definitions

The terms ‘value system,’ ‘value,’ and ‘valuation’ have a range of meanings in different disciplines. In this paper, we provide a practical synthesis of these concepts in order to address the issue of valuation of ecosystem services. We want to be clear about how we use these terms throughout our analysis. ‘Value systems’ refer to intrapsychic constellations of norms and precepts that guide human judgment and action. They refer to the normative and moral frameworks people use to assign importance and necessity to their beliefs and actions. Because ‘value systems’ frame how people assign rights to things and activities, they also imply practical objectives and actions. We use the term ‘value’ to mean the contribution of an action or object to user-specified goals, objectives or conditions (Costanza, 2000). A specific value of that action or object is tightly coupled

---

* Corresponding author. Tel.: +1-412-648-7602
E-mail addresses: eofarbr@vms.cis.pitt.edu (S.C. Farber),
costza@cbl.umces.edu (R. Costanza), mwilson@cbl.umces.edu
(M.A. Wilson).

1 Tel.: +1-410-326-7263.
2 Tel.: +1-301-405-8952.
3 As of 09/01/2002, this author can be reached at the Gund
Institute for Ecological Economics, University of Vermont,
School of Natural Resources, George D. Aiken Center,
Burlington VT 05405–0088, USA.
with a user’s value system because the latter determines the relative importance of an action or object to others within the perceived world. We define ‘valuation’ as the process of expressing a value for a particular action or object. In the current context, ecosystem valuation represents the process of expressing a value for ecosystem goods or services (i.e. biodiversity, flood protection, recreational opportunity), thereby providing the opportunity for scientific observation and measurement.

The distinction between intrinsic and instrumental value is an important one (Goulder et al., 1997). On the one hand, some individuals might maintain a value system in which ecosystems or species have intrinsic rights to a healthful, sustaining condition that is on a par with human rights to satisfaction. The value of any action or object is measured by its contribution to maintaining the health and integrity of an ecosystem or species, per se, irrespective of human satisfaction. Some interpret Leopold and Aldo (1949) land ethic as constituting an intrinsic value system, where something is ‘right when it tends to preserve the integrity, stability and beauty of the biotic community. It is wrong when it tends otherwise.’ On the other hand, instrumental values reflect the difference that something makes to satisfaction of human preferences. Instrumental values, such as economic values, are fundamentally anthropocentric in nature. Policies toward the environment will always tend to be based on a mix of intrinsic and instrumental value systems. In this paper, we deal with both.

2. Economic concepts of value

The history of economic thought is replete with struggles to establish the meaning of value; what is it and how is it measured. Aristotle first distinguished between value in use and value in exchange. The paradox of use versus exchange value remained unresolved until the 16th century (Schumpeter and Joseph, 1978). The diamond–water paradox observed that while water has infinite or indefinite value, being necessary for life, its exchange value is low; yet unessential diamonds bear a high exchange value. Following this observation, there was widespread recognition of the distinction between exchange value and use value of goods. Galiani defined value to mean a relation of subjective equivalence between a quantity of one commodity and a quantity of another. He noted that this value depends on Utility and Scarcity (utilita et rarita) (Schumpeter and Joseph, 1978). Two hundred years later, Adam Smith distinguished between exchange value and use value of goods by citing the diamond-water paradox, but used it to dismiss use value as a basis for exchange value. Smith formulated a cost of production theory of value, whereby wages, profit and rent are the three original sources of exchange value. In his famous beaver–deer example he suggested a labor theory of exchange value: if it takes twice the labor to kill a beaver than to kill a deer, one beaver will sell for as much as two deer. He also suggested a labor-disutility theory of exchange value, noting that goods exchange based upon the unpleasantness of the labor required to bring the goods to market. However, it is significant to note that Smith limited his labor theory to ‘that early and rude state of society which precedes both the accumulation of stock and the appropriation of land’. In other words, when labor is the only scarce factor, goods will exchange based upon the ratio of labor use (Schumpeter and Joseph, 1978).

In addition to formulating his hypothesis regarding the origins of exchange value, Smith sought to establish a unit of measure of value, or what he termed the real measure or real price of a good. He proposed that ‘labour alone...never varying in its own value, is alone the ultimate and real standard’ of the values of all commodities. Hence labor could be a numeraire, and it had special properties of invariant value (Schumpeter and Joseph, 1978).

Ricardo also sought an invariant unit of measure for value. He felt that there was no commodity, including labor, whose exchange value could serve as an invariant standard to measure the variation in exchange values of other commodities. And it was not possible to add up commodities to measure national wealth or production with only exchange ratios. According to Ricardo,
this measure must be invariant to changes in relative factor rewards, i.e. capital versus labor, and be a commodity whose capital and labor use did not vary over time, i.e. no technological change. He proposed that both wheat and gold possessed these properties (Blaug and Mark, 1968). While not creating value they could measure value.

While Ricardo had several followers, including J.S. Mill and Marx, labor theories of value and the pursuit of an invariant standard of value waned in the late 19th century. This was partially in response to the logic of the utilitarians, such as Menger, Gossen, Jevons and Walras, who argued that exchange value was based on both utility and scarcity (Blaug and Mark, 1968). Sraffa, a noted Ricardian scholar, sought to resurrect the classical pursuit of a theory of value independent of demand or value in use. In his book, Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory, Sraffa (1960) established conditions under which exchange ratios between commodities can be determined based on their use in production; i.e. a set of commodity prices that would exhaust the total product. These exchange ratios were not based on any optimality or marginality conditions. Instead, Sraffa divided commodities into basic (goods which entered into all production processes) and non-basic, and showed that an invariant standard of value would be a combination of basic commodities reflecting average input proportions in production. This contrived ‘commodity’ would then be usable as a measure of national wealth or income.³

The ‘marginal’ revolution in value theory originated with the confluence of several related streams of economic thought in the 20th century. Menger proposed there were different categories of wants or desires, such as food, shelter, clothing, etc., that could be ordered in terms of their subjective importance. Within each category, there is an ordered sequence of desires for successive increments of each good. He postulated that the intensity of desire for one additional unit declines with successive units of the good (Blaug and Mark, 1968). Replacing the term ‘desire for one additional unit’ with the term ‘Marginal Utility,’ we thus have the economic principle of diminishing marginal utility.

The idea that people have different, but ordered, categories of wants or desires raises the critical issue of whether trade-offs exist between categories. If individuals ‘weight’ categories, it implies a trade-off. At one extreme, categories may be lexicographically ordered, like words in a dictionary. One level of want must be satisfied before a lower level becomes relevant in the process of valuation. There are no trade-offs between levels of wants. For example, the need for caloric intake is likely superior to that of recreational pleasure — no number of recreational opportunities will likely substitute for an insufficient diet. In the lexicographic case, individuals would use their monetary resources hierarchically, satisfying higher order wants and needs first. When a higher order want or need is at risk, the individual would take resources away from lower level ones until higher level needs were satisfied. Lexicographic preferences do not mean monetary valuation is impossible, as individuals would still be able to state how much of their resources they would be willing to sacrifice for a good or service; but it may be all their resources if a high level need is at risk.

More problematic for valuation are instances where basic needs cannot be satisfied by the resources at an individual’s disposal — i.e. time or money. Similar to Menger, Ekins et al. (1992) suggested the universality of basic human needs, including subsistence, affection, protection, understanding, leisure, identity, and freedom. Although one can imagine needs like affection being ‘purchasable’ with money, or ‘freedom’ being purchasable by migration, many of these needs may not be satisfied by money or time because individuals simply may not consider them to be purchasable by money or time. Thus, not only is it possible that trade-offs between needs will not be possible, but some needs may not be reducible to money or time.

³ While accepting Sraffa (1960) mathematical proof, some reviewers (Harrod, 1961; Reder, 1961) noted that the exchange values would not be independent of demand as Sraffa claimed. It was further noted that Sraffa’s did not constitute a price theory in the sense of establishing the process of price determination.
Lancaster and Kelvin (1971) introduced the concept of consumption technology, whereby consumers consider characteristics of goods. For example, food may be evaluated on caloric, protein or vitamin content. Different foods are substitutable depending on the composition of their characteristics. People allocate their budget across characteristics, purchasing goods that are efficient sources of desired characteristics. The technological inability to substitute characteristics may restrict the margins on which environmental goods and services can be valued. For example, while health may be valued, and individuals would be willing to pay for it, the proper mix of calories, protein and vitamins may make marginal increases or decrements in one of these characteristics either very highly valued or of very low value.

Building on this insight, multi-attribute utility theory formalizes the utility-generating technology by proposing that total utility is a function of the characteristics of goods or services. A simple example would be where utility, \( U \), from food consumption is a linear function of the caloric, \( C \), protein, \( P \), and vitamin, \( V \), content:

\[
U = aC + bP + cV. \tag{1}
\]

Here, the parameters \( a \), \( b \), and \( c \) reflect the weighting of three factors in determining utility for food consumption. When utilities are measurable in monetary willingness to pay (WTP) or willingness to accept (WTA) compensation, these parameters represent the marginal monetary value of each characteristic. This logic forms the basis for hedonic pricing models of valuation, discussed below, whereby the value of market goods, say a house, depends upon the characteristics of the house and its location, as well as surrounding environmental amenities or disamenities.

Gossen proposed that in order to maximize satisfaction from a good, such as labor or money, an individual must allocate that good across different uses to equate their marginal utilities in each use (Blaug and Mark, 1968). Hence marginal utility would provide a basis for explaining exchange value. If we treat things such as iron, cement, fertilizer, natural agents and labor as incomplete consumable goods, the marginal utility of the goods they produce can be used to explain their exchange value. This logic established a full theory of value. It also demonstrated that exchange values could be based on use value. While the diamond–water paradox had been solved many times, the classical economists, such as Smith and Ricardo, could not resolve it using their labor theories of value. It was resolved only by recognizing the importance of utility and scarcity in determining exchange values, and the role of margins in value determination.

While the classical theorists sought a standard physical commodity unit for measuring exchange value, neoclassical theorists did not need such a commodity. As value was assumed to be determined by utility on the margin, and consumers were assumed to allocate money optimally across uses, the marginal utility of money was the same for an individual in all its uses. Money thus became the standard unit of measure.

The significance of the marginal utility theory of value to the evolving concept of ecosystem service valuation is that it can be used to measure use values, not just exchange values, in monetary units. The general optimization model of labor/leisure and consumption/saving given time and wealth constraints would yield equivalencies of goods for money, goods for time, and time for money. Time or money can thus be used as a standard of measure of use value; how much time or money will a person willingly sacrifice to obtain commodity \( X \)? In sum, as the pursuit of an economic theory of value traversed the broad metaphysical terrain of economic thought, the answer appears to have been found in the concept of value in use.\(^4\)

The utility-based values of goods and services are reflected in people’s WTP to attain them, or their WTA compensation to forego them. WTP and WTA become measures of these values. They may be based on small marginal changes in the availability of these goods and services, or on larger changes including their complete absence or

\(^4\) Since the marginal utility of a good depends upon how much the person possesses, we would expect a difference depending upon whether the person is asked how much they would sacrifice to obtain \( X \) or how much would they accept in compensation to forego \( X \) (see Hicks, 1939).
presence. These valuations are reflected in Fig. 1. Let the curve D represent the WTP for each unit of the good or service, \( T \), for an individual or group. This is a ‘Marginal’ WTP. The ‘Total’ WTP for \( T_0 \) units of \( T \) is the aggregated areas \( A + B \). Area \( A \) may be very large for goods or services that have some utility threshold where the good becomes increasingly valuable as it becomes scarcer. This is true for many ecological goods and services, such as life support goods like oxygen and water; the ‘Marginal’ value is finite but the ‘Total’ value is indeterminate. This is the distinction that lies behind the diamond-water paradox noted above.

Exchange-based values are reflected in the prices, \( P \), at which the goods or services are exchanged. When supply is \( T_0 \), and the item is sold competitively, a price \( P \) is determined which clears the market. These prices also reflect the ‘Marginal’ valuations placed on available quantities around \( T_0 \). So prices reflect ‘Marginal’ values when there are markets for the goods or services.\(^5\) The ‘Total’ exchange value of \( T_0 \) is \( P \times T_0 \). This is an observable market value when there are markets to observe. But when there are no such markets, \( P \) must be determined indirectly, and \( P \times T_0 \) would represent a pseudo-market value. This would be the ‘Total’ exchange value of the good if there were a market with an available supply of \( T_0 \).

Measures of economic value are designed to reflect the difference that something makes to satisfaction of human preferences. If something is attainable only at a cost, then the difference it makes to satisfy preferences is the difference between its utility and the cost of attaining it. Formal concepts of Compensating and Equivalent Variations are used to reflect this difference (Varian and Hal, 1992). For example, suppose in Fig. 1 that \( T_0 \) is available at a cost of \( P \). Under these terms of availability, the welfare difference made by \( T_0 \) is area \( A \). The ‘Marginal’ value that alterations in availability make to welfare would be reflected by changes in \( A \). Using timber from trees as an example, suppose timber is harvested at a cost of \( P \) per unit of timber. The value of trees, per se, would be represented by area \( A \), which is less than \( A + B \).

Thus conceived, the basic notion of value that guides economic thought is inherently anthropocentric, or instrumental. While value can generally mean the contribution to a goal, objective, desired condition, etc., the mental model used by economists is that value is based on want satisfaction, pleasure or utility goals. Things have value insofar as they propel individuals toward meeting pleasure and need objectives. Values of objects in the environment can be considered on the margin, as well as on the whole; i.e. the value of one additional tree versus the value of all trees. While value relates to the utility of a thing, the actual measurement of value requires some objective measure of the degree to which the thing improves pleasure, well-being, and happiness.

In a finite world, the resources people have available to meet their personal objectives are limited. Economists have thus developed an extensive theory of how people behave in the presence of constraints on feasible activities (Varian and Hal, 1992). The working hypothesis is that people make decisions in order to optimize satisfaction, pleasure or utility. This optimization always takes place in the presence of constraints, such as income, wealth, time, resource supply, etc. Optimization thus yields a deterministic set of possible decisions in most real-world situations—when constraints change, so do the decisions.

\(^5\) Unfortunately, this is not the case for many unmarketed ecological goods and services—techniques that economists have developed for assessing the ‘Marginal’ values of goods are outlined below in Section 5.
The essence of this perspective is that the economic world works largely deterministically, moving from one equilibrium to another in relatively stable fashion, and responds to changes in constraints in a predictable fashion. The determination of equilibrium is a resultant of conflicting forces, such as supply and demand, or unlimited wants and limited means. While there are instances of instability, disequilibria and indeterminism, these are treated as exceptions rather than the rule.

Since individuals can be observed making choices between objects in the marketplace while operating within the limits of income and time, economists have developed measures of value as imputations from these observed choices. While monetary measures of value are not the only possible yardstick, they are convenient since many choices involve the use of money. Hence, if you are observed to pay $10 for a bottle of wine, the imputation is that you value wine to be at least $10, and are willing to make a trade-off of $10 worth of other things to obtain that bottle. The money itself has no intrinsic value, but represents other things you could have purchased. Time is often considered another yardstick of value; if you spend 2 h golfing, the imputation is that you value the golf experience to be worth more than 2 h spent in other activities. Value is thus a resultant of the expressed tastes and preferences of persons, and the limited means with which objects can be pursued. As a result, the scarcer the object of desire is, the greater its value will be on the margin.

Importantly, the ‘technologies’ of pleasure and production allow for some substitution between things. A variety of goods can induce pleasure and are thus treated conceptually as utility substitutes. A bear may substitute for an elk in consumption, hunting, and in a wildlife viewing experience even though bears and elk are not substitutes in terms of ecosystem function. On the production side, inputs are also considered to be substitutable for one another. Machines and technology can substitute for people and natural inputs. Clearly, economists recognize that the relations between goods and services are often more complicated than this. For malnourished people, sugar is no technological substitute for protein, even though they both provide calories. As discussed earlier, preferences may be lexicographic—some things are more important than others, and cannot be substituted for lower level wants or needs. On the production side, no number of lumbermen is a substitute for timber when there is no timber. Production may require certain inputs, but at the same time there may be substitutability between others. As Krutilla and John (1967) suggests, there may be close substitutes for conventional natural resources, such as timber and coal, but not for natural ecological systems.

The neoclassical perspective also assumes that tastes and preferences are fixed and given, and that fundamental economic ‘problem’ consists of optimally satisfying those preferences. Tastes and preferences usually do not change rapidly and, in the short run (i.e. 1–2 years), this basic economic assumption is probably not too bad. In the longer run, however, it does not make sense to assume tastes and preferences are fixed. People’s preferences do change over longer time frames as the existence of a robust advertising industry attests. This observation is important because sustainability is an inherently long-run concept and ecosystem services are expected to continue into the far future. This fact is very disturbing for many economists because it takes away the easy definition of what is optimal. If tastes and preferences are fixed and given, then we can adopt a stance of ‘consumer sovereignty’ and just give people what they want. We do not have to know or care why they want it; we just have to satisfy their preferences efficiently.

However, if preferences change over time and under the influence of education, advertising, changing cultural assumptions, and variations in abundance and scarcity, etc., we need a different criterion for what is ‘optimal’. Moreover, we have to figure out how preferences change, how they relate to this new criterion, and how they can, or should, be changed to satisfy the new criterion (Norton et al., 1998). One alternative for the new criterion is sustainability itself, or more completely a set of criteria: sustainable scale (size of the economic subsystem), fair distribution, and efficient allocation (Daly, 1992). This set of crite-
ria implies a two-tiered decision process (Page 1977; Daly and Cobb, 1989; Norton et al., 1998) of first coming to a social consensus on a sustainable scale and fair distribution and, second, using the marketplace and other social institutions like education and advertising to implement these decisions. This might be called 'community sovereignty' as opposed to 'consumer sovereignty.' It makes most economists very uncomfortable to stray from consumer sovereignty because it raises the question: if tastes and preferences can change, then who is going to decide how to change them? There is a real danger that a totalitarian government might be employed to manipulate preferences to conform to the desires of a select elite rather than the individuals in society.

Here, two points need to be kept in mind: (1) preferences are already being manipulated every day; and (2) we can just as easily apply open democratic principles to the problem as hidden or totalitarian principles in deciding how to manipulate preferences. Viewed in this light, the aforementioned question is transformed: do we want preferences to be manipulated unconsciously, either by a dictatorial government or by big business acting through advertising? Or do we want to formulate preferences consciously based on social dialogue and consensus with a higher goal in mind? Either way, we believe that this issue can no longer be avoided, and is one that will best be handled using open democratic principles and innovative thinking. Which leads us back to the role of individual preferences in determining value. If individual preferences change in response to education, advertising, and peer pressure then value cannot solely originate with individual preferences. Values ultimately originate from within the constellation of shared goals to which a society aspires—value systems—as well as the availability of 'production technologies' that transform things into satisfaction of human needs.

In addition to income and education, time places constraints on value creation. Constraints of time and intertemporal substitutabilities create temporal implications for value. Economists presume that a present time preference exists due to limited time horizons and concerns for uncertainty in the future (Fisher, 1930). This means individuals will discount values of things in the future in comparison to the same things in the present. If I have an equal endowment of apples now and a year from now, I would place a greater value on having an apple now than on having an apple 1 year from now. The ability to convert things to money in the presence of positive financial interest rates will, therefore, result in the 'optimizing individual' discounting things in the future.

In contrast to economists’ traditional assumptions of positive time preferences, or positive discount rates, psychologists suggest time preference is more complicated. For example, Loewenstein and Prelec (1991) find that in some circumstances people behave as if they have negative time preference, preferring more in the future to more now. The authors suggest this is due to dread, the anticipation of savoring better conditions in the future, and the aversion to loss. However, this negative time preference may not be operative when the time period is ambiguous. The implications of such experimental results for discounting in environmental policy settings are not clear, but they do raise serious questions about the standard practice of discounting future environmental benefits (Clark, 1973).

3. Ecological concepts of value

‘Value’ is a term that most ecologists and other natural scientists would prefer not to use at all, except perhaps in its common usage as a reference to the magnitude of a number—e.g. ‘the value of parameter b is 9.32’ Using the definition of value provided earlier, ecosystems and non-human species are presumed not to be pursuing any conscious goals, and, therefore, they do not have a ‘value system’. Likewise, one cannot talk about ‘value’ as the degree to which an item contributes to achieving a goal in this context since there is no conscious goal being pursued. Nevertheless, some concepts of value are important in the natural sciences, and are in fact quite commonly used, and we try briefly to elucidate them here.

If one limits the concept of value to the degree to which an item contributes to an objective or
condition in a system, then we can see how natural scientists use the concept of value all the time to talk about causal relationships between different parts of a system. For example, one could talk about the value of particular tree species in controlling soil erosion in a high slope area, or the value of fires in recycling nutrients in a forest.

There are other ways in which the concept of ‘value’ is used in the natural sciences. For example, a core organizing principle of biology is evolution by natural selection. Evolution in natural systems has three components: (1) generation of genetic variation by random mutations or sexual recombination; (2) natural selection by relative reproductive success; (3) transmission via information stored in the genes. While this process does not require conscious, goal-directed behavior on the part of any of its participants, one can still think of the overall process as being ‘goal-directed’. The ‘goal’ of ‘survival’ is embedded in the objective function of natural selection. While the process occurs without consciousness of this goal, species as a whole can be observed to behave ‘as if’ they were pursuing the goal of survival. Thus, one often hears evolutionary biologists talk about the ‘survival value’ of particular traits in organisms. Natural selection models, which maximize the fitness of species, are not only testable, they bear close similarities to economic utility maximization models (Low, 2000).

Beyond this, the idea of ‘co-evolution’ among a whole group of interacting species (Ehrlich and Raven, 1964) raises the possibility that one species is ‘valuable’ to the survival of another species. Extending this logic to the co-evolution of humans and other species, we can talk of the ‘value’ of natural ecosystems and their components in terms of their contribution to human survival.

Ecologists and physical scientists have also proposed an ‘energy theory of value’, either to complement or replace the standard neoclassical theory of value (Odum, 1971, 1983; Slesser, 1973; Gilliland, 1975; Costanza, 1980; Cleveland et al., 1984; Hall et al., 1992). It is based on thermodynamic principles where solar energy is considered to be the only primary input to the global ecosystem. This theory of value represents a return to the classical ideas of Ricardo and Sraffa (see above), but with some important distinctions. The classical economists recognized that if they could identify a ‘primary’ input to the production process then they could explain exchange values based on production relationships. The problem was that neither labor nor any other single commodity was really ‘primary’.

The classical economists were writing before the physics of thermodynamics had been fully developed. Energy—or, more correctly, ‘free’ or ‘available’ energy—has special characteristics which satisfy the criteria for a ‘primary’ input: (1) Energy is ubiquitous. (2) It is a property of all of the commodities produced in economic and ecological systems. (3) While other commodities can provide alternative sources for the energy required to drive systems, the essential property of energy cannot be substituted for. Available energy is thus the only ‘basic’ commodity and is ultimately the only ‘scarce’ factor of production, thereby satisfying the criteria for a production-based theory that can explain exchange values.

Energy-based concepts of value must follow the basic principles of energy conversion. The first law of thermodynamics tells us that energy and matter are conserved. But, this law essentially refers to heat energy and mechanical work (raw energy or the bomb calorimeter energy). The ability to do work is related to the degree of organization or order of a thing relative to its environment, not its raw energy content. Heat must be organized as a temperature gradient between a high temperature source and a low temperature sink in order for useful work to be done. In a similar fashion, complex manufactured goods like cars have an ability to do work that is not related to their raw energy content. The second law of thermodynamics tells us that useful energy (organization) always dissipates (entropy or disorder always increases) within a closed system. In order to maintain organized structures (like an economy) one must constantly add organized, low entropy energy from outside the system.

Estimating total ‘energy’ consumption for an economy is not a straightforward matter because not all fuels are of the same quality—i.e. they vary in their available energy, degree of organization, or ability to do work. Electricity, for exam-
ple, is more versatile and cleaner in end use than petroleum, and it also costs more energy to produce. In an oil-fired power plant it takes from 3–5 kcal of oil to produce each kcal of electricity. Thus, adding up the raw heat equivalents of the various forms of fuel consumed by an economy without accounting for fuel quality can radically distort the picture, especially if the mix of fuel types is changing over time.

An energy theory of value posits that, at least at the global scale, free or available energy from the sun (plus past solar energy stored as fossil fuels and residual heat from the earth’s core) are the only ‘primary’ inputs to the system. Labor, manufactured capital, and natural capital are ‘intermediate inputs’. Thus, one could base a theory of value on the use in production of available energy that avoids the problems the classical economists encountered when trying to explain exchange values in economic systems. There have been a few attempts to empirically test this theory using both timeseries data and cross-sectional data. Studies that have tried to adjust for fuel quality have shown a very close relationship between ‘available energy’ consumption and economic output. Cleveland et al. (1984) and more recently Kaufmann (1992) have shown that almost all of the changes in E/GNP (or E/GDP) ratios in the US and OECD countries can be explained by changes in fuel quality and the percent of personal consumption expenditures (PCE) spent directly on fuel. The latter effect is due to the fact that PCE is a component of GNP and spending more on fuel directly will raise GNP without changing real economic output. Fig. 2 is an example of the explanatory power of this relationship for the US economy from 1932 to 1987. Much of the apparent gain in energy efficiency (decreasing E/GNP ratio) is due to shifts to higher quality fuels (like natural gas and primary electricity) from lower quality ones (like coal). Renewable energy sources are generally lower quality and shifts to them may cause significant increases in the E/GNP ratio.

Another way of looking at the relationship between available energy and economic output uses cross-sectional rather than time-series data. This avoids some of the problems associated with changes in fuel mix and distortions in GNP. For example, Costanza (1980), Costanza and Herendeen (1984) used an 87-sector input–output model of the US economy for 1963, 1967, and 1973, modified to include households and government as endogenous sectors (to include labor and government energy costs) to investigate the relationship between direct and indirect energy consumption (embodied energy) and dollar value of output. They found that dollar value of sector output was highly correlated \( R^2 = 0.85–0.98 \) with embodied energy, though not with direct energy consumption or with embodied energy calculated excluding labor and government energy costs. Thus, if one makes some necessary adjustments to estimates of energy consumption in order to better assess ‘available energy’, it appears that the empirical link between available energy and economic value is rather strong.

Some neoclassical economists have criticized the energy theory of value as an attempt to define value independent of consumer preferences (see Heuttner, 1976). This criticism is axiomatic as the stated purpose was to establish a biophysical theory of value not completely determined by social preferences. The energy theory of value overcomes some of the problems with production-
based theories of value encountered by the classical economists discussed earlier and does a reasonable job of explaining exchange values empirically in the few cases where it has been tested. Despite the controversy and ongoing debate about the validity of an energy theory of value (Brown and Herendeen, 1996), it seems to be the only reasonably successful attempt to operationalize a general biophysical theory of value (see Patterson, this volume).

4. Ecological thresholds, uncertainty, and economic value

Ecosystems can be highly non-linear within certain regions, and changes can be dramatic or irreversible (see Limburg et al., this volume). The availability of ecosystem services may be dramatically altered at these non-linear points for only minor changes in ecosystem conditions. A valuable service provided to humans by naturally functioning ecosystems is their avoidance of adverse threshold conditions, or what Ciriacy-Wantrup (1963) referred to as ‘Critical Zones’ for resource conservation. For example, trees in a forested ecosystem provide a hydrologic service of moderating water flows into streams during peak storm events. As Fig. 3 below shows, let us suppose there is a relationship between the density of trees in a landscape and physical severity of downstream flooding. At tree densities exceeding the ‘Critical Threshold,’ marginal changes in density can be evaluated using measures such as expected increases in flood damages. Under this marginal regime, there is a substitute for nature’s services, flood protection or property replacement. Below the critical threshold, however, flood severity increases substantially as tree density diminishes. Economic values change substantially for slight alterations in ecosystem conditions because human lives and communities may be at substantial risk. Under these conditions, traditional monetary measures of value may not be able to adequately capture the impact of severe floods. Traditional valuation methods may not be acceptable as measures of the values of trees in proximity of the ‘Critical Threshold’.

Due to the probabilistic nature of storm events, human society may wish to maintain tree densities well in excess of the critical threshold, say at $T^*$.

There would be a welfare loss if tree densities fell below $T^*$, and this loss would be attributable to both the marginal increase in flood severity and to the fact that now the system is closer to a catastrophe. There would be an insurance premium that society would pay to avoid such a dramatic change in ecological states. Additional trees would have value both for their role in reducing expected flood damages, and as insurance for avoiding a natural catastrophe.

The example above illustrates that ecosystem service value has both efficiency and sustainability components. In the linear, marginal region, where the actual states of the economic and ecological systems are not dramatically altered, the values of changing tree densities are rationally based on efficiency goals; in this case avoiding having to repair flood damages. In the non-linear, non-marginal region, however, the value of trees is a sustainability value, as they protect the economic and ecological systems from collapse. Sustainability values may be more important than efficiency values around and below threshold limits. In short, sustainability values may be lexicographically superior to efficiency values.

The example in Fig. 3 has the property of reversibility. Even when tree densities fall below the Critical Threshold level and place society at high risk, planting more trees reverses the exposure to risk. This may not be the case with some ecosystem conditions. For example, reductions in tree densities below the Critical Threshold may alter landscape conditions for a long period of...
time even after tree densities have been increased to pre-threshold conditions.

In Fig. 4, the ecological–economic system moves along path cd rather than ba once the threshold of irreversibility has been violated. This irreversibility would likely increase the value society would pay to avoid the threshold compared with conditions of relatively easy reversibility. The insurance value would include not only a premium to avoid a catastrophe, but an option value to avoid the irreversibility of flooding (Arrow and Fisher, 1974).

The recent wildfire in Los Alamos, New Mexico, in the summer of 2000, provides a dramatic, tragic example of the catastrophes and irreversibilities associated with being near critical thresholds. The fire was started as a controlled burn of several hundred acres by the US National Park Service. Years of improper forest management, such as natural fire suppression and grazing of understory vegetation created a circumstance in which a minor change, the small controlled burn, had disastrous consequences, destroying 300 homes and temporarily displacing 30,000 people. To make matters worse, the destruction of groundcover over nearly 50,000 acres will likely permanently alter soil conditions as erosion will be very severe. The former forest system may never be replicated. This situation is similar to conditions illustrated in Fig. 4.

Another example may be the value of trees in a landscape. In Fig. 5, alterations in tree densities above the ‘Critical Threshold’ level only marginally change the visual appeal of the landscape. However, below this critical threshold the landscape is no longer a forest; the state of nature is altered substantially. Changes in tree densities above the critical threshold can be valued on the margin using traditional economic valuation techniques. However, suppose the forest is a critical visual element to a community, or the loss of forest has dire impacts on the state of the local

![Fig. 4. Flood protection values of trees with ecosystem irreversibilities.](image)

![Fig. 5. The value of trees for visual appeal.](image)
economy or social fabric. Changes in tree densities below the critical threshold may not be meaningfully valued using traditional techniques. Furthermore, the same type of insurance value as in the case of flooding will give a premium to remaining above the critical threshold. Given there are probabilistic events such as storms and infestations, the community may wish to keep densities above $T^\ast$. Increases in density in the region around the critical threshold have value both for improving the visual appeal of the landscape as well as providing insurance that tree densities will not fall below this threshold.

In all of the above examples, critical thresholds in ecosystem structure or function do not necessarily imply economic thresholds for values. For example, if flooding impacts on communities were never severe or people were not highly dependent on the existence of a forest, marginal economic valuation methods would be appropriate across the whole range of tree densities. This is in spite of the fact that there may be thresholds of densities at which ecosystem structures and functions are substantially altered. The natural world may be in a non-linear, non-marginal condition, but the economic world remains a smooth one where substitutes readily mitigate significant ecosystem change. Of course, the opposite may be true also—i.e. gradual changes in natural conditions may lead to non-linear changes in economic conditions. For example, water quality may gradually fall below certain standards and a lake is ‘suddenly’ closed to swimmers.

Critical thresholds where ecological conditions and dynamics are uncertain require valuation under uncertainty. Uncertainty may range from knowing the probabilities of conditions and their values, to only being able to identify the conditions but not their probabilities. There are several methods for dealing with such valuation dilemmas. For example, suppose an ecosystem under State A would provide $200 in services, but in State B would provide $0 in services. If the probability of each state occurring is 0.5, the expected value of the ecosystem services is $100. An example would be the storm protection value of an acre of coastal barrier islands under a hurricane (State A) or no hurricane (State B).

Under these conditions, the valuation of ecosystem services is not quite so simple. Individuals may be averse to losses and a gain may be weighted more heavily than a gain of comparable magnitude. Given this, what would be the WTP to preserve the barrier islands; what are they worth? The answer depends on whether the decision maker is risk averse. There are two uncontrollable states: Storm versus No Storm, each occurring with a 0.5 probability. There are two ecological conditions: Barrier Island and No Barrier Island.

Table 1

<table>
<thead>
<tr>
<th>Barrier island conditions</th>
<th>No barrier island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storm occurs</td>
<td>$I - C + $200</td>
</tr>
<tr>
<td>No storm</td>
<td>$I$</td>
</tr>
<tr>
<td>Expected value</td>
<td>$I + 0.5 (S200 - C)$</td>
</tr>
<tr>
<td>Worst case</td>
<td>$I - C + $200</td>
</tr>
</tbody>
</table>

Using expected values, the value of the barrier islands is $100. However, the WTP to maintain the islands is given by:

$$0.5U(I - C + S200 - WTP) + 0.5U(I - WTP) = 0.5U(I - C) + 0.5U(I)$$  

or an amount such that the expected utility, net of WTP, of maintaining the barrier islands just equals the expected utility without the islands. It can be shown that under conditions of risk aversion, where the utility function is concave, WTP would be greater than the expected value of the loss, $100 in this case, but less than the full damage of $200.\textsuperscript{6} The excess of WTP over the expected value of the loss is the ‘premium’ that

\textsuperscript{6}When WTP = $100, rearranging terms shows that $U(I - C + 100) - U(I - C) > U(I) - U(I - 100)$ for any concave utility function.
risk averters would pay rather than risk a full loss. This example can be generalized in several ways. First, if the barrier islands have some additional value, such as recreational or aesthetic enjoyment, the value of the islands measured by WTP would be additive to the storm protection values. Second, when altering ecological conditions increases the probability that a loss will occur, risk averting individuals should be willing to pay something to avoid the increase in probability of loss. This WTP would reflect what the ecosystem is worth insofar as insuring against crossing thresholds and encountering adverse irreversible conditions.

When uncertainty consists of not knowing the probabilities of various ecological states, e.g. of a hurricane, the above matrix can be used to illustrate valuation under this pure risk situation. If society is risk averse, a useful decision rule is to assume the worst will occur, and seek to minimize the worst-case scenario. For example, maintaining the barrier islands results in a worst-case scenario under a hurricane of $I - C +$200; while the worst case if islands are not maintained is $I - C$. The implied value of the barrier islands for planning purposes is $200 under this risk averse decision rule, as society would be willing to pay up to $200 to maintain these islands.

In this example, the ability to estimate storm damage cost savings provides guidance to valuing the resource as well as developing a decision rule. A simple decision rule would be to maintain and conserve an ecosystem service when the cost of doing so is not too great. This is the ‘safe minimum standard’ proposed by Ciriacy-Wantrup (1963), and elaborated on by others (Bishop, 1978). Under this standard, conservation practices avoid Ciriacy–Wantrup’s ‘Critical Zone’ of dramatic, irreversible change in ecosystems. For example, soil conservation would avoid gulleys or maintain a maximum acceptable erosion rate; forest conservation would establish a maximum deforestation rate; rangelands conservation would maintain a minimum level of plant material after grazing; or species conservation would establish a minimum breeding stock or habitat condition. These standards are ecologically based, not economic; although violating them may be prudent if the economic costs are too high.

5. Conflicts between economic and ecological values

We also recognize that economic and ecological measures of value may at times be at odds with one another. As humans are only one of many species in an ecosystem, the values they place on ecosystem functions, structures and processes may differ significantly from the values of those ecosystem characteristics to species or the maintenance (health) of the ecosystem itself. The intrinsic values of natural system features and processes within the natural system itself may possess different abundance and functional value properties than their corresponding economic values. Diminishing returns and utility would suggest some economic saturation in the demands for particular ecosystem services and conditions. For example, the marginal economic value for additional sunlight may be zero or possibly negative—skin cancer from excessive sunlight, excessive heat, etc.

The differences between ecological and economic values relate to the relative abundance of ecosystem services within naturally functioning ecosystems and economies. Clearly, a service can be more abundant or scarce in one than another. While it is likely that specific ecosystem structures and processes have some functional role in an ecosystem, and, therefore, have ‘value,’ they may not have direct or indirect value in market economies. There may be instances where an ecosystem is so isolated from human economic activity that what happens in it is irrelevant to human activity, even when all possible spatial and temporal connections are considered—i.e. only the intrinsic value remains. Of course, as humans continue to increasingly inhabit the planet, these instances become increasingly rare. As our understanding of connections between and within ecosystems expands, we find more and more instances of significant implications for human beings. These changing conditions in knowledge make it increasingly incumbent upon us to avoid the quick dismissal of isolated or presumably
economically irrelevant ecosystems or their properties as irrelevant to human welfare.

6. Economic valuation methods

The exchange value of ecosystem services is the trading ratios for those services. When services are directly tradable in normal markets, the price is the exchange value. The exchange-based, welfare value of a natural good or service is its market price net of the cost of bringing that service to market. For example, the exchange-based value of timber to society is its 'stumpage rate,' which is the market price of timber net of harvest and time allocated management costs. Exchange-based valuation is relatively simple, as trades exist from which to measure values.

Market prices reflect the valuation of goods and services, but only on the margin. For example, the price of a board foot of timber reflects what another board foot is worth to buyers. It does not reflect the whole value of all timber used by the buyers. You may pay only $2 per board foot and that is all you would be willing to pay for the last of, say, the 5000 board feet you buy. But you may be willing to pay considerably more than $10,000 for the opportunity to buy all 5000 board feet. Of course, the timber reflects only a portion of the full social value of a tree, which also provides an array of services such as soil amendment and stabilization, water storage and flood control, species habitat, aesthetics, climate control, etc. In limited cases, markets for environmental services have been formed that tend to reflect the valuations of those services (Chichilnisky and Heal, 1998).

While exchange value requires markets or observable trades, the social value of services is much more broad and difficult to measure. These social values are what have captured the attention of environmental and resource economists. They have developed a number of techniques for valuing ecosystem services (Freeman, 1993; Kopp and Smith, 1993). The underlying concepts for social values that economists have developed are what a society would be willing and able to pay for a service, WTP, or what it would be willing to accept to forgo that service, WTA. The two valuation concepts may differ substantially in practice (Hannemann, 1991).

The economic valuation methodology essentially constructs WTP for a service; or constructs the adequate compensation for a service loss, representing WTA. Suppose the service is flood control provided by a wetland. Suppose damages from flooding were $1 million. Society would then be willing to pay $100,000 to reduce the probability of flooding by 10% if the society, as a whole, is risk neutral. Suppose the wetlands reduce flooding probabilities by 20%. When wetlands services are free, society receives $200,000 million in services for nothing. In principle, the owner of a wetland providing such a service could capture up to this amount of social value if there was a capture mechanism. Markets for resource services provide capture mechanisms. They work relatively well for 'private' goods, where owners can deny access to the service if payments are not made and if making access available to one person essentially makes it unavailable to others. Raw materials and food production are good examples of these 'private' goods or services.

Many ecosystem services do not qualify for market trading because they are not ‘private’ in nature. For example, flood protection services of wetlands or trees, once made available to one person may indirectly become available to all. Wetlands and forest owners could not capture all the potential social WTP for this service.

When there are no explicit markets for services, we must resort to more indirect means of assessing economic values. A variety of valuation techniques can be used to establish the WTP or WTA for these services. There are six major ecosystem service economic valuation techniques when market valuations do not adequately capture social value:

- **Avoided Cost (AC):** services allow society to avoid costs that would have been incurred in the absence of those services; flood control avoids property damages or waste treatment by wetlands avoids health costs.

- **Replacement Cost (RC):** services could be replaced with man-made systems; natural waste treatment can be replaced with costly treatment systems.
• Factor Income (FI): services provide for the enhancement of incomes; water quality improvements increase commercial fisheries catch and incomes of fishermen.
• Travel Cost (TC): service demand may require travel, whose costs can reflect the implied value of the service; recreation areas attract distant visitors whose value placed on that area must be at least what they were willing to pay to travel to it.
• Hedonic Pricing (HP): service demand may be reflected in the prices people will pay for associated goods; housing prices at beaches exceed prices of inland homes.
• Contingent Valuation (CV): service demand may be elicited by posing hypothetical scenarios that involve some valuation of alternatives; people would be willing to pay for increased fish catch or deer bag.

Each of these methods has its strengths and weaknesses. Also, each service has an appropriate set of valuation techniques. Some services may require that several techniques be used jointly. For example, the recreational value of an ecosystem will include not only the value that visiting recreationists place on the site (TC), but the increased incomes associated with site use (FI). Alexander et al. (1998) have suggested an extreme FI for valuing global ecosystem services, measuring the rents that a hypothetical monopolistic owner of nature’s services could charge the world’s economy. For example, an extreme measure of rent from all natural system services would be the difference between global GDP and a global subsistence income. The paper by de Groot, Wilson and Boumans (this volume) discusses the appropriate techniques for valuing different ecosystem services.

Some valuation techniques, while intuitively appealing, may misrepresent WTP or WTA valuation concepts in certain circumstances. This is especially a problem when using Replacement Cost (RC) methods. There may be circumstances when the social benefits that may be lost when ecosystem services are unavailable are less than the cost of replacement of those services; or when the benefits gained from enhanced services are less than alternative means of providing those services. For example, the Avoided Cost of illness under an ecosystem enhancement, such as wetlands treatment of waste, may be less than the cost of comparable waste treatment facilities. In this case, Avoided Cost is a more appropriate measure of value than Replacement Cost. The Replacement Cost measure of value of the world’s coral reefs may far exceed the measure of benefits.

7. The challenge of aggregating economic values

The traditional procedure of economic valuation is to establish individual-based values using one of the methods described in Section 5 above. Isolated individual values are then aggregated to represent a socially-relevant unit—a community, a state, a nation, or the entire planet. This is appropriate when the services provided are purely individually enjoyed, as is the case for ‘private’ goods and services that are not shared and where there are no substantial positive or negative (externality) impacts of one person’s use on another. This is also the case for ‘public’ goods where enjoyment remains individual-based without externality impacts. An example would be the recreational enjoyment of an uncongested forest.

Isolated, individual-based valuation and aggregation are not appropriate, however, in instances where group values may hinge on group interactions, where preference formation is partially a social process, where shared knowledge is important, and where items valued have substantial interpersonal or social implications. Valuing a forest for timber, or even recreation is appropriately an individual-based process. However, other values of the forest may be more communal, are not well-defined in preference functions, or have substantial interpersonal impacts. For example, the value of forests to a community whose social system, folklore, etc. are intimately dependent on them is more than the sum of independent personal values.

One approach to ecosystem service valuation that has gained increasing recognition in the literature is small group deliberation (Jacobs, 1997; Blamey and James, 1999; Coote and Lenaghan, 1997). Derived from political theory, this evolving
set of techniques are founded on principles of deliberative democracy and the assumption that public decision making should result, not from the aggregation of separately measured individual preferences, but from a process of open public debate (Fishkin, 1991; Dryzek, 1987; Habermas, 1984). Thus, the application of a participatory democracy approach to environmental issues establishes two validity criteria that set it apart from traditional non-market valuation approaches: decentralized forms of environmental policy formulation and the direct involvement of non-experts in small decision-making groups (see Wilson and Howarth, this volume).

The basic idea is that small groups of citizens can be brought together to deliberate about the social value of public goods and that the ‘consensus’ values derived in this open forum can then be used to guide environmental public policy (Jacobs, 1997). In this manner, discursive methods such as citizens’ juries (Coote and Lenaghan, 1997), consensus conferences (James and Blamey, 1999), and deliberative CV techniques (Sagoff, 1998) have increasingly been proposed and used in North America, Europe, and Australia to inform environmental decision making. One assumption common to all these techniques is that deliberative bodies of citizens can render informed judgments about environmental goods not simply in terms of their own personal utility, but also for society as a whole. The purpose of deliberation is to ‘reach agreement on what should be done by or on the behalf of society as a whole’ (Jacobs, 1997). In sum, open discourse is assumed to perform a ‘corrective function’ when each citizen alone has incomplete information, but acting together with others can piece together a more complete picture of true social value for ecosystem goods and services.

For example, we might consider the recently proposed deliberative, or ‘group’ CV technique (Sagoff, 1998; Jacobs, 1997). While there is a long tradition of group research in CV, the goal of such research has generally been to use focus groups to increase the content-validity of hypothetical scenarios and diagnose potential problems that individual respondents may have with the payment vehicle (Mitchell and Carson, 1989). With a group CV, on the other hand, the explicit goal would be to derive a group-consensus value for the ecological good or service in question. The valuation exercise is, therefore, conducted in a manner similar to a conventional CV survey—using hypothetical scenarios and realistic payment vehicles—with the key difference being that value elicitation is not done through private questioning but through group discussion and consensus building. Thus, the deliberative CV approach treats small group deliberation not as a diagnostic tool, but as an explicit mechanism for value elicitation.

8. Conclusions

The concepts of ‘value’, ‘value system’, and ‘valuation’ have many meanings and interpretations and a long history in several disciplines. We have provided a survey of some of these meanings as they relate to the issue of ecosystem service valuation to serve as background and introduction to the remaining papers in this special issue. There is clearly not one ‘correct’ set of concepts or techniques to address this important issue. Rather, there is a need for conceptual pluralism and thinking ‘outside the box.’ That is what the remaining papers in this special issue attempt to do. While they break some new ground and address the issues in interesting new ways, it is clear that much additional work remains to be done. After a long and interesting history, the issue of ‘value’ is now going through a period of development that should help us to make better, and more sustainable, decisions, not only as individuals, but also as groups, communities, and as stewards of the entire planet.

Acknowledgements

This work was conducted as part of the Working Group on the Value of the World’s Ecosystem


References


